# ESTIMATION AND TESTING PROCEDURES FOR THE PARAMETERS OF THE NEGATIVE BINOMIAL DISTRIBUTION


By

LINDA JEAN WILLSON

Bachelor of Science
West Texas State University
Canyon, Texas
1974

Master of Science
West Texas State University
Canyon, Texas
1976



Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1981

My son, if thou wilt receive my words, and hide my commandments with thee, so that thou incline thine ear unto wisdom, and apply thine heart to understanding; yea, if thou criest after knowledge, and liftest up thy voice for understanding; if thou seekest her as silver, and searchest for her as for hidden treasures; then shalt thou understand the fear of the Lord, and find the knowledge of God. For the Lord giveth wisdom; out of His mouth cometh knowledge and understanding. He layeth up sound wisdom for the righteous; He is a shield to those who walk uprightly. He keepeth the paths of justice, and preserveth the way of His saints. Then shalt thou understand righteousness, and justice, and equity; yea, every good path. When wisdom entereth into thine heart, and knowledge is pleasant unto thy soul, discretion shall preserve thee, understanding shall keep thee.

Proverbs 2: 1 - 11

ESTIMATION AND TESTING PROCEDURES FOR

THE PARAMETERS OF THE NEGATIVE

BINOMIAL DISTRIBUTION

Thesis Approved:

J. Leroy Folks
Thesis Adviser

Ronald W. McNew

David L. Weeks

Jerry H. Young

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

It is difficult for me to fully express the deep sense of appreciation I have for each of you who have encouraged me in the pursuit of this degree and who now share a part in it with me. The faculty and staff have produced an atmosphere that makes learning enjoyable. Perhaps I could have graduated without the help of my fellow graduate students, but I would have learned less. They have made the bad times bearable and the good times better. The people in the community, especially those in my church, have provided a balance in my life that has helped me keep education in its proper place. To all these people, I can only say thank you.

There are a few who should be mentioned by name. My committee is composed of Dr. J. Leroy Folks, Dr. Ronald McNew, Dr. David Weeks and Dr. Jerry Young. I chose these men because of the high regard I have for their abilities, and this respect has increased with time. I also consider each to be a friend of mine. Dr. Folks, my major adviser, exhibited the "patience of Job" as I struggled to learn the true meaning of research. He skillfully guided my work, and when "the going got tough," he would assure me that there was a "light at the end of the tunnel." Dr. McNew was always ready to listen when the solution of a problem was eluding me. He kept reminding me that all of my odd pieces of work were not "junk" but simply "unpolished gems." Dr. Weeks has a special talent for expressing his love and concern for each graduate student. During my time at OSU, he has become my confidant, and he has taught me much about statistics and life. Dr. Young gave me the first opportunity I had to apply statistics. He suggested the general area of research, and his enthusiasm for the work was contagious. I am very proud of each member of my committee, and I thank them for the time and effort they have contributed to helping me both in and out of the classroom.

I would like to express my appreciation to Dr. Nitis Mukhopadhyay for the many helpful discussions we had during the course of my research. My thanks also go to Dr. Robert D. Morrison for the learning experience he gave me. He taught me many things about statistics and working with people.

I want to acknowledge the financial support I have received during the past four years from the Department of Statistics, the Department of Entomology, and the Oklahoma Agricultural Experiment Station. I have received invaluable work experience from each of these. Most of the computer funds used during the course of this study were secured by Dr. Young through the Department of Entomology. I am very grateful for this funding because, without it, many of the Monte Carlo studies would not have been possible. I will be indebted forever to Kate Mellina for the excellent job she has done typing this thesis. She showed great concern for the project, and overlooked my errors and pickiness.

I thank my parents, Mr. and Mrs. Marvin B. Cornette, for instilling in me a desire to learn and beginning me on the road of education. I appreciate my son, John Jayret, for the many hours of enjoyment he has provided, the help he has given, and the sacrifices he has made. My last note of thanks goes to the person to whom I owe the most, my husband and best friend, Johnny. Without his support, I would not have been able to begin this degree. He has been a never-ending source of encouragement,

patience, and love, and when all else failed, he would help me see the humorous side of a problem.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Page

# CHAPTER I

## INTRODUCTION

Although discussed by Pascal and Fermat, the negative binomial distribution was first formulated and published by Montmort (24) in 1714.

In 1907, Student (36) encountered the negative binomial while studying the distributions of yeast cells counted with a haemocytometer. He reasoned that if the liquid in which the cells were suspended was properly mixed, then a given particle had an equal chance of falling on any unit area of the haemocytometer. Thus he was working with the binomial distribution and the fact that the probability a binomial random variable X assumes a value x is equal to the (x + 1)st term in the expansion of $(p + q)^k$ where p, q, and k > 0 and p + q = 1; that is,

$$P(X = x) = \binom{k}{x} p^x q^{k-x} , \quad x = 0, 1, 2, \ldots, k .$$

Student estimated p, q, and k from the first two sample moments. In two of his four series, the second moment exceeded the mean, resulting in negative estimates of p and k. Nevertheless, these "negative" binomials fit his data well. He noted that this may have occurred due to a tendency of the yeast cells "to stick together in groups which was not altogether abolished even by vigorous shaking" (p. 357).

There were several other cases that appeared in the literature during the early 1900's where estimation of the binomial parameters resulted in negative values of p and k. This phenomenon was explained to some extent by arguing that for small p and large k the variability of the estimators would cause some negative estimates to be observed. Whitaker (38) investigated the validity of this claim. In addition to Student's work, she reviewed that of Mortara (25) who dealt with deaths due to chronic alcoholism and that of Bortkewitsch (7) who studied suicides of children in Prussia, suicides of women in German states, accidental deaths in trade societies, and deaths from the kick of a horse in Prussian army corps. In view of the estimated errors associated with the various estimates of p and the frequency of negative estimates, Whitaker found it highly unlikely that all negative estimates of p and k could be explained by variability. She, therefore, suggested that a new interpretation was needed for the negative binomial, $(q - p)^{-k}$, where $p > 0$, $k > 0$, and $q = 1 + p$. By expansion of this expression, we find that the probability the negative binomial random variable X will assume the values 0, 1, 2, ... is

$$P(X = x) = \binom{k + x - 1}{k - 1} \frac{p^x}{q^{x+k}} \quad . \tag{1.1}$$

In 1920, Greenwood and Yule (15) developed an accident proneness model. They began by considering a Poisson random variable X; hence,

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad , \quad x = 0, 1, 2, \ldots$$

where $\lambda > 0$ represents the expected number of events for an individual in the population. If the value of $\lambda$ differs from one individual to the next, and if $\lambda$ is distributed according to the cumulative probability function $F(\lambda)$, the probability of observing x events in the total population is given by

$$\int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} \, dF(\lambda) \quad . \tag{1.2}$$

Greenwood and Yule refer to (1.2) as a compound Poisson distribution. Upon considering various forms of $F(\lambda)$, they discovered the negative binomial resulted when $\lambda$ varied according to a gamma distribution (12, pp. 391-394).

The negative binomial distribution was also derived in 1923 by Eggenberger and Pólya (11) as a limiting case of an urn problem in the following manner. Suppose an urn contains N balls of which Np are red and Nq are white ($p + q = 1$). There are n successive drawings of a ball made from the urn with replacement, and $N\delta$ balls of the color last drawn are added to the urn after each drawing. Let X denote the number of red balls in n successive drawings. Then the probability X assumes the value x is given by

$$P(X=x) = \binom{n}{x} \frac{p(p+\delta)(p+2\delta)\ldots(p+(x-1)\delta)q(q+\delta)(q+2\delta)\ldots(q+[n-x+1]\delta)}{1(1+\delta)(1+2\delta)\ldots(1+[n-1]\delta)} . \tag{1.3}$$

If we let $n \to \infty$, $p \to 0$, and $\delta \to 0$, while keeping $np = \lambda$ and $n\delta = \eta$ constant, then (1.3) becomes

$$P(X = x) = \frac{k(k+1)\ldots(k+x-1)}{x!} \left(\frac{\eta}{1+\eta}\right)^{x} \left(\frac{1}{1+\eta}\right)^{\frac{\lambda}{\eta}} .$$

On setting $\frac{\lambda}{\eta} = k$, it is apparent from (1.1) that this is a negative binomial distribution (16, pp. 1392-1393).

The negative binomial has come to be applied in many fields including accident statistics, population counts, psychological data, and communications.

Our main interest in this distribution arises from our involvement with entomological problems. We hope to develop methods, with a strong statistical foundation, that researchers will be able to use easily. Some of the procedures presented in this thesis have been field-tested on cotton insects.

The mean is of primary interest in entomology, as it is in most applications. Anscombe (2) reparameterized the distribution in the 1940's, using $\mu = kp$ and $k$. A random variable $X$ is then distributed as a negative binomial random variable if the probability mass function is given by

$$P(X = x) = \binom{k + x - 1}{k - 1}\left(\frac{\mu}{\mu + k}\right)^{x}\left(\frac{k}{\mu + k}\right)^{k}, \quad x = 0, 1, 2, \ldots$$
$$= 0 \qquad , \quad \text{otherwise} \tag{1.4}$$

where $\mu$ and $k$ are positive. Here the mean is $\mu$ and the variance is $\mu + \frac{\mu^2}{k}$. It is evident that the variance both exceeds the mean and is a quadratic function of the mean. Although various equivalent forms exist in the literature, we shall consider Anscombe's which has become almost standard.

In this thesis, our main objective will be to investigate some problems of statistical inference related to the negative binomial distribution. The negative binomial may be viewed as a one-parameter distribution where either $\mu$ or $k$ is unknown or a two-parameter distribution where both $\mu$ and $k$ are unknown. The two-parameter distribution is difficult to work with, and many simplifications result when we can assume that we know $k$. However, when $k$ is the unknown parameter, inference is only slightly improved over the two-parameter problem. We do not believe that this last case arises very often in practice. Thus we shall consider the cases where $\mu$ is the unknown parameter and where both $\mu$ and $k$ are unknown.

In Chapter II, we shall present some of the properties of the negative binomial distribution and discuss earlier research in our areas of interest. Sequential procedures for estimation of the mean in the one-parameter case will be developed in Chapter III, and a nonparametric, sequential approach to estimation of the mean which is applicable to the two-parameter negative binomial distribution is studied in Chapter IV.

Chapters V and VI focus on inference related to the second parameter k. In Chapter V, a multistage procedure for estimating k is presented. A proposed, fixed-sample-size estimation and testing procedure for a value of k common to several populations is developed and compared to a standard one in Chapter VI.

## CHAPTER II

## SOME STATISTICAL PROPERTIES AND PRIOR RESEARCH ON THE NEGATIVE BINOMIAL DISTRIBUTION

In this chapter, we shall review some of the properties and previous work that has been done on the negative binomial distribution.

### Properties of the Negative Binomial Distribution

It is interesting to note how each parameter affects the shape of the negative binomial. In Figures 1 - 4, $\mu$ and k have a similar effect on the shape of the distribution. When one parameter is held fixed, an increase in the second one results in a shift of the distribution to the right. As the value of the fixed parameter increases, the shape of the distribution is more dramatically affected by changes in the second parameter. The similarity in the behavior of the parameters is further evidenced when we note that both $\mu$ and k must approach infinity for there to be no skewness or kurtosis.

The probability generating function is

$$\phi_X(t) = \left(1 + \frac{\mu(1 - t)}{k}\right)^{-k}, \quad |t| < \frac{k + \mu}{k} \quad . \tag{2.1}$$

Since $\phi_X(t)$ is well-determined for $|t| < \frac{k + \mu}{k}$ and $\frac{k + \mu}{k} > 1$, all moments do exist. From (2.1), we can determine that the r-th factorial about 0 is

$$E[X^{(r)}] = (r - 1)! \binom{k + r - 1}{k} \frac{\mu^r}{k^{r-1}} \, .$$

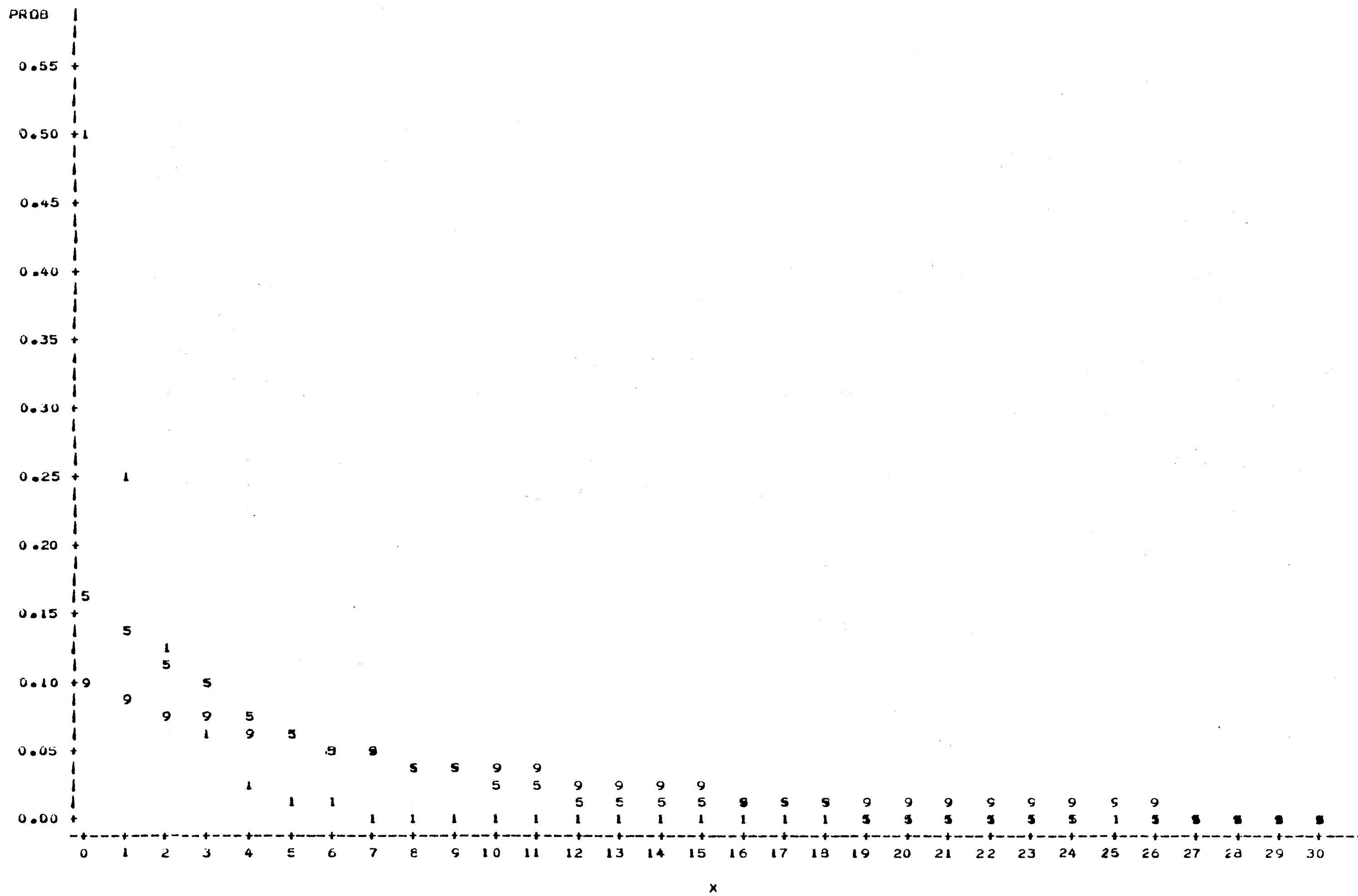



Figure 1. Three Probability Mass Functions With a Fixed k = 1 and Differing Values of μ: μ = 1 (1), μ = 5 (5), and μ = 9(9)

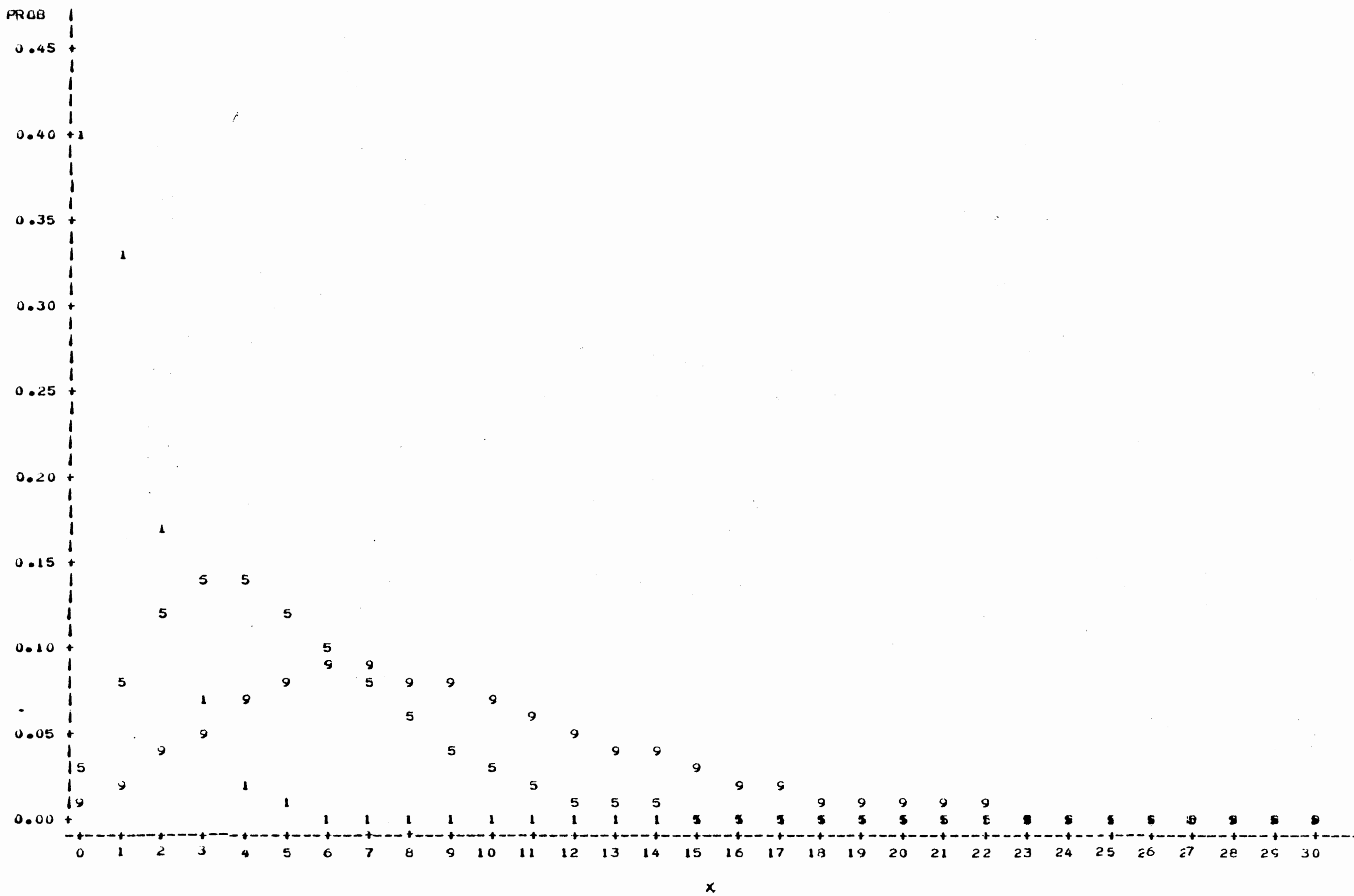



Figure 2. Three Probability Mass Functions With a Fixed k = 5 and Differing Values of μ: μ = 1 (1), μ = 5 (5), and μ = 9 (9)

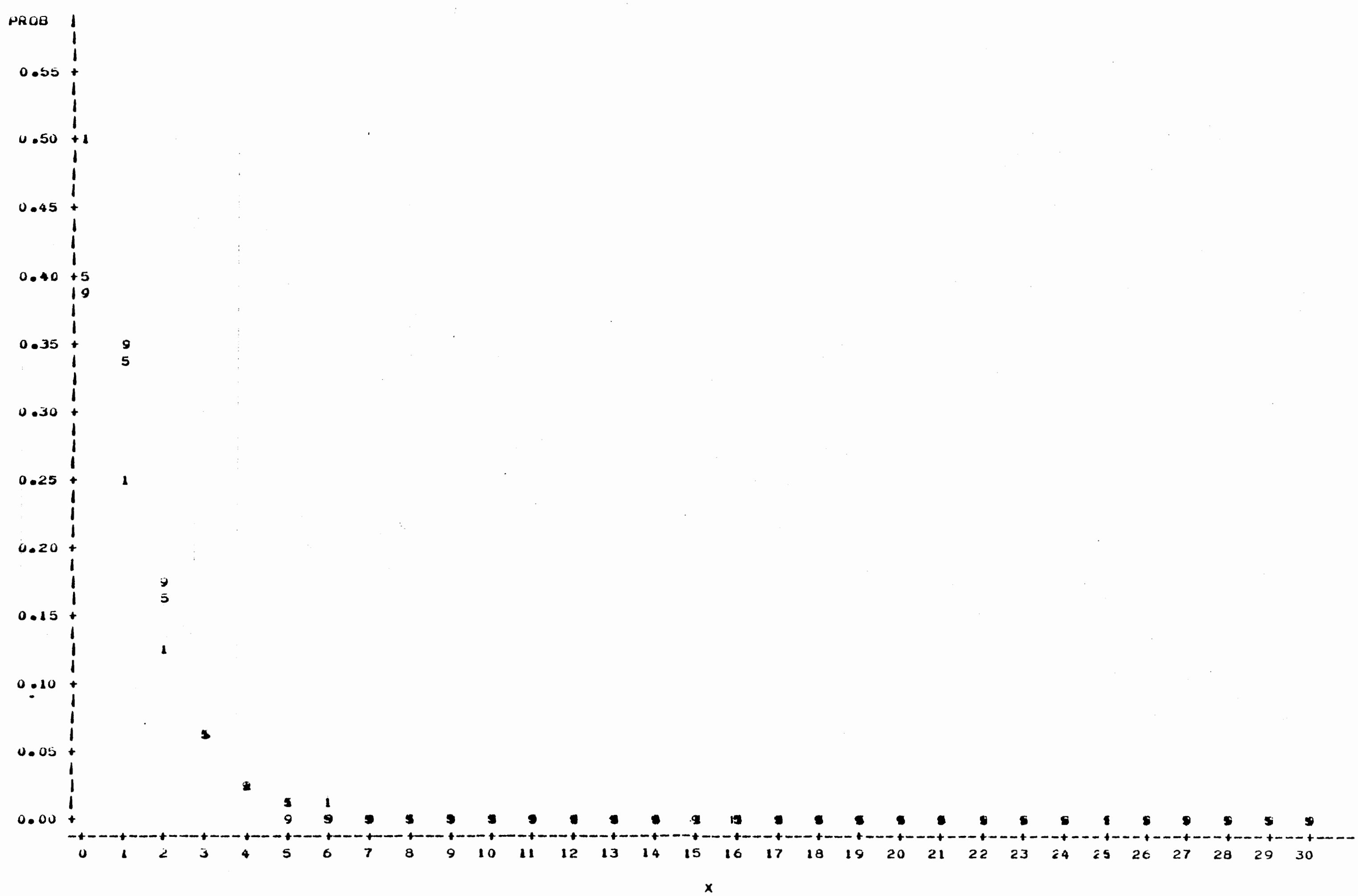



Figure 3. Three Probability Mass Functions With a Fixed $\mu = 1$ and Differing Values of k: k = 1 (1), k = 5 (5), and k = 9 (9)

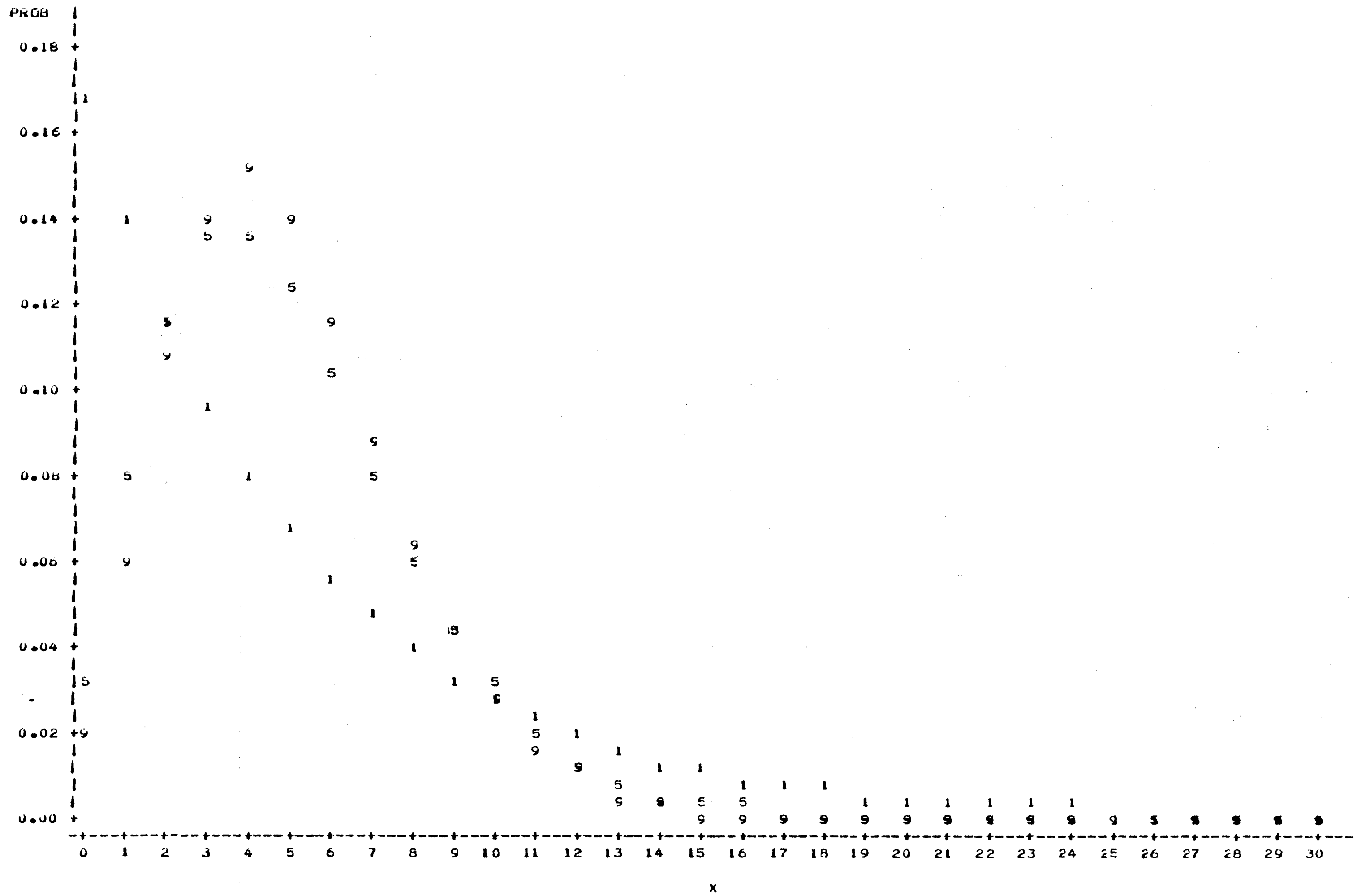



Figure 4. Three Probability Mass Functions With a Fixed $\mu = 5$ and Differing Values of k: $k = 1$ (1), $k = 5$ (5), and $k = 9$ (9)

The moment generating function is given by

$$M_X(t) = \left(1 + \frac{\mu(1 - e^t)}{k}\right)^{-k}, \quad t < \ln \frac{k + \mu}{\mu} \quad . \tag{2.2}$$

Using (2.2), it may be shown that the sum of n independent, identically-distributed negative binomial random variables with parameters $\mu$ and k, $NB(\mu, k)$, is also distributed as $NB(n\mu, nk)$.

The negative binomial distribution with parameter $\mu$ belongs to the exponential family. It is easily proven that $\sum_{i=1}^{n} x_i$ is a complete, sufficient statistic. Thus, given any sample of fixed size n, $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the minimum variance unbiased estimator of $\mu$.

## Previous Sequential Estimation of $\mu$ for the One-Parameter Negative Binomial

Suppose we are interested in estimating the mean of a one-parameter negative binomial distribution with a prescribed level of precision. Let $n^*$ be the minimum fixed sample size required to obtain the desired precision of the estimate. We shall refer to $n^*$ as the optimal fixed sample size. Usually, $n^*$ depends on unknown parameters and is therefore unknown. As a result, samples of random size N are used in the estimation process. When working with sequential or multistage procedures, we shall need some notation that indicates the randomness of the sample size. Define $x_i$ to be the i-th observation from the population of interest. Denote the sum of the first n observations by $T_n$ and their average by $\overline{X}_n$. Also, let

$$s_n^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \overline{X}_n)^2$$

and

$$u_n^2 = \frac{1}{n}\left(\sum_{i=1}^{n} (x_i - \overline{X}_n)^2 + 1\right).$$

Oakland (30) was the first to apply Wald's sequential probability ratio test to tests of hypotheses about $\mu$ for a negative binomial distribution. In doing so, he made two assumptions. The first was that the value of k would be the same under both hypotheses, and the second was that either k would be known or a precise estimate of k would be available. Morris (26) employed Oakland's procedures after estimating k using methods developed by Anscombe (2).

Sequential point estimation of the mean attempts to estimate the parameter $\mu$ with a prescribed degree of precision. The choice of an appropriate procedure depends upon the method used to measure the precision. One approach attempts to estimate $\mu$ with a specified coefficient of variation of the mean, C. Estimation of $\mu$ within a proportion p of the mean with confidence $1 - \alpha$ is the goal of some procedures, and a third method is designed to estimate the mean within d units with confidence $1 - \alpha$.

In 1969, Kuno (19) presented two sequential procedures for the estimation of the mean when the variance is of the form

$$a\mu + b\mu^2$$

where a and b are constants. Assuming a and b are known, he estimates the mean using $\overline{X}_n$ and the variance by $a\overline{X}_n + b\overline{X}_n^2$ .

The first of Kuno's procedures is designed to estimate $\mu$ with a specified coefficient of variation of the mean, C. Observations are taken until the first time the estimated $CV(\overline{X})$ is less than or equal to the desired C; that is, until

$$\sqrt{\frac{(a\overline{X}_n + b\overline{X}_n^2)/n}{\overline{X}_n^2}} \leq C \quad . \tag{2.3}$$

Then the estimate of the mean is taken to be $\overline{X}_N$ . The advantage of this rule is that it may be rewritten: Stop the first time that

$$T_n \geq \frac{a}{C^2 - \frac{b}{n}} \ .$$

Thus all major computations may be done before taking the sample, and only the total needs to be calculated in order to determine whether or not to stop. In a subsequent paper, Kuno (20) presented an asymptotic expression of the bias associated with this method of estimation and found it to be

$$\frac{aC^2\mu}{a + b\mu} \ .$$

With this exception, the statistical properties of the process have not been investigated. Therefore, from Kuno's work, we do not know if the procedure actually achieves the desired coefficient of variation of $\overline{X}$.

Allen, Gonzalez, and Gokhale (1) used this process and determined the values of a and b by regression. From their data on Heliothis zea, the bollworm, they had 49 estimated mean-variance pairs each based on seventy-two five-plant samples taken in five different fields over a three-year period. They determined the least squares estimates of a and b in the equation,

$$s^2 = a\overline{X} + b\overline{X}^2 \ .$$

Once a and b were estimated, Allen, Gonzalez, and Gokhale employed Kuno's procedure.

Kuno's second sequential procedure attempts to estimate the mean with a specified standard deviation of $\overline{X}$, $d_0$ . This problem is closely related to estimation of $\mu$ within d units with confidence $1 - \alpha$. He suggests taking observations until

$$\sqrt{\frac{a\bar{X}_n + b\bar{X}_n^2}{n}} \leq d_0 \ .$$

The stopping criterion can be written so that all major calculations may be completed before drawing the sample. Thus we would stop the first time

$$T_n \leq \frac{n(\sqrt{a^2 + 4nbd_0^2} - a)}{2b} \ .$$

The asymptotic bias for this process is

$$B = -\frac{d_0^2}{\mu} \cdot \frac{a + 2b\mu}{a + b\mu} \ .$$

Again no investigation into whether or not this sequential procedure actually attains the desired level of precision has been made.

Later, Binns (5) presented a method of estimating the mean of a one-parameter negative binomial distribution within a proportion p of $\mu$ with confidence $1 - \alpha$. The optimal fixed sample size required to achieve this goal is

$$n^* = \left(\frac{z\sigma}{p\mu}\right)^2$$

where z is the $1 - \frac{\alpha}{2}$ fractile of the standard normal distribution. He also uses $\bar{X}_n$ as a preliminary estimate of $\mu$ and $\bar{X}_n + \frac{\bar{X}_n^2}{k}$ as the estimate of the variance. Thus he recommends adding observations to the sample sequentially until

$$n \geq \left(\bar{X}_n + \frac{\bar{X}_n^2}{k}\right)\left(\frac{z}{p\bar{X}}\right)^2 \ . \tag{2.4}$$

Letting $a = \frac{z^2}{p}$ and employing a finite population correction factor on nk, we may rewrite (2.4) as stop when

$$nk > a^2 + \frac{1}{2} \quad \text{and} \quad T_n \geq a^2 + \frac{a^4}{nk - \frac{1}{2} - a^2} \,. \tag{2.5}$$

Denoting by $(t, r_t)$ the point of intersection of the sample path and the stopping boundary $(N - 1 < t \leq N)$, $\mu$ is then estimated by

$$\hat{\mu} = \frac{kr_t}{tk - \frac{1}{2}} \quad .$$

Note that $\mu$ is basically the estimated value of $\overline{X}$ when the sample path crosses the stopping boundary with the addition of the finite population correction factor on tk.

Binns showed that for large a the distribution of the estimate is approximately log normal with mean log $\mu$ and variance $\frac{1}{a^2}$ . He approximated the average sample size and the variance of the sample size and investigated the adequacy of these approximations using Monte Carlo methods. The effect of imprecise knowledge of k was also studied. Although Binns showed that the estimate had some nice properties, he did not determine whether or not the procedure achieved its goal of estimating $\mu$ within $p\mu$ with confidence $1 - \alpha$.

## Research Related to Nonparametric Sequential Estimation Procedures for $\mu$

Other sequential procedures are available when the distribution of the population is unspecified. Chow and Robbins (9) studied the properties of a method designed to estimate $\mu$ within a specified distance d with confidence $1 - \alpha$. The stopping rule is of the form

$$N = \min\left(n \geq n_0 \ (\geq 1): \quad n \geq \left(\frac{u_n z}{d}\right)^2\right).$$

Taking $\overline{X}_N$ as the estimate of $\mu$, they proved

$$\lim_{d \to 0} \frac{N}{n^*} = 1 \quad \text{a.s.}$$

$$\lim_{d \to 0} P(|\overline{X}_N - \mu| \leq d) = 1 - \alpha$$

$$\lim_{d \to 0} \frac{E(N)}{n^*} = 1 \quad .$$

Sproule (33) extended the work of Chow and Robbins to cover the means of U-statistics.

Nadás (29) developed a process for estimating $\mu$ within a proportion p of the mean with confidence $1 - \alpha$. Using $\hat{\mu} = \overline{X}_N$ and the stopping rule

$$N = \min\left(n \geq n_0 \ (\geq 1): \left(\frac{u_n}{\overline{X}_n}\right)^2 \leq n\left(\frac{p}{z}\right)^2\right)$$

where again z is the $1 - \frac{\alpha}{2}$ fractile of the standard normal distribution, he proved

$$\lim_{p \to 0} \frac{N}{n^*} = 1 \quad \text{a.s.}$$

$$\lim_{p \to 0} P(|\overline{X}_N - \mu| \leq p\mu) = 1 - \alpha$$

$$\lim_{p \to 0} \frac{E(N)}{n^*} = 1 \quad .$$

Sequential point estimation of the mean when the distribution is normal with parameters $\mu$ and $\sigma$ was studied by Starr (34). He considered a loss structure,

$$L_N = A|\overline{X}_N - \mu|^s + N^t$$

where A, s, $t > 0$. Introducing the term risk efficiency, he showed that the ratio of the expected loss (or risk) associated with the sequential process to the risk associated with the optimal fixed sample size tends to one as $\sigma \to \infty$.

Mukhopadhyay (28) developed a nonparametric, sequential procedure for estimating the mean when the loss function is

$$L_N = A(\bar{X}_N - \mu)^2 + cN$$

where A and c are known positive constants, c being the cost per observation. Two assumptions were needed to prove the risk efficiency of the process: (1) $0 < \sigma_0 < \sigma < \infty$ for a known $\sigma_0$ and (2) $E(|X|^{2+\delta}) < \infty$ for some $\delta > 0$.

Ghosh and Mukhopadhyay (14) further developed this last sequential method so that the only distributional assumption is the finiteness of the eighth moment. This process was extended to cover U-statistics by Sen and Ghosh (32).

## Research on the Estimation of k

When considering the two-parameter negative binomial distribution, the estimation of k poses a problem. The method of moments estimator (MME) of k for a fixed sample size n is

$$\hat{k} = \frac{\bar{X}_n^2}{s_n^2 - \bar{X}_n} .$$

Since $k > 0$, this estimate is not reasonable when the estimate of the mean exceeds that of the variance. Haldane (17) derived the maximum likelihood estimators (MLE). Using $m_j$ to denote the number of times a j was observed in a sample of size n, the MLE of k is the root of the following equation in $\hat{k}$:

$$n \ \ln\left(1 + \frac{\bar{X}_n}{\hat{k}}\right) = \sum_{j=1}^{\infty} m_j \left(\frac{1}{\hat{k}} + \frac{1}{\hat{k} + 1} + \ldots \frac{1}{\hat{k} + j - 1}\right) .$$

Anscombe (3) hypothesized that there is only one positive finite root when $s_n^2 > \overline{X}_n$ and none otherwise. Fisher (13) compared the asymptotic efficiency of the MME and MLE. Other fixed sample size estimates are available but seldom used.

Bowman and Shenton (8) presented formulas for computing the bias of the method of moments and maximum likelihood estimators of k. Tables of these biases were given for order $\frac{1}{n}$ and $\frac{1}{n^2}$. Using computer simulation to draw samples of size 50, 100 and 200, Pieters, Gates, Matis, and Sterling (31) compared methods of estimating k. Their conclusion was that there appeared to be little difference in the biases under the method of moments and maximum likelihood. However, both the method of moments and maximum likelihood were superior to the other estimation procedures considered.

## Estimation and Testing for a Common k

Since insect counts are often fit well by the negative binomial distribution, there have been a number of attempts to give a meaningful ecological interpretation of the parameters $\mu$ and k. $\mu$ may be defined as the average density of insects in the area of interest. The definition of k has been more elusive.

Anscombe (2) stated that k depends on the intrinsic power of a species to reproduce itself, while $\mu$ depends on external factors. This has led some researchers to search for an inherent value of k associated with various species.

Waters (37) suggested that k measures the aggregation of insects. Following his logic, small values of k indicate extreme aggregation whereas the distribution of counts tends to be purely random as k approaches infinity.

The idea of mean crowding and its relationship to k was explored by Lloyd (22). Mean crowding, $\mu^*$, is defined to be the mean number per individual of other individuals in the same sampling unit. If the underlying distribution is negative binomial, then $\frac{\mu^*}{\mu}$, the ratio of mean crowding to the population mean, is $1 + \frac{1}{k}$. Assuming that this ratio is constant for a given species of insect, we would then have another interpretation for k.

There is no doubt that all of these are valid to some extent. However, there is a tendency to extend these interpretations to inferences about the spatial distribution of insects. We believe that this is totally incorrect.

Since there has been a vast amount of work devoted to the meaning of k, it is important to be able to test for the equality of k's from populations with differing means and to estimate that common k if it exists. Let $n_i$, $s_i$, and $\overline{X}_i$ denote the sample size, estimated standard deviation, and estimated mean, respectively, of the i-th population, $i = 1, 2, \ldots, t$. Anscombe (3) presented some methods of estimating a k common to several populations with differing means. The most popular approach was to choose $\hat{k}_c$ so that the sum or weighted sum $s_i^2 - \overline{X}_i - \frac{\overline{X}_i^2}{\hat{k}_c}$ was zero. In order to minimize the variance of $\hat{k}_c$, he suggested using a weight for the i-th population of

$$w_i = \frac{n_i - 1}{(\overline{X}_i + \hat{k}_c)^2} .$$

In 1958, Bliss and Owen (6) recommended computing

$$x_i' = \overline{X}_i - \frac{s_i^2}{n_i}$$

and

$$y_i' = s_i^2 - \overline{X}_i$$

for each of the t populations. The regression line of $y'$ on $x'$ passes through the origin and has slope $\frac{1}{\hat{k}_c}$. In order to increase the precision of the estimate, each population may be weighted inversely to the variance. Hence,

$$w_i = \frac{.5(n_i - 1)k^4}{k(k + 1) - {}^{(2k - 1)}/_{n_i} - {}^{3}/_{n_i^2}} \; \frac{1}{\overline{X}_i^2(\overline{X}_i + k)^2} \; .$$

After iteratively performing a weighted regression until the last two estimates of the common k differ by a negligible amount, a chi-square test for equality of the k's from each of the t populations can then be conducted. This testing and estimation procedure is the one most commonly employed at present.

# CHAPTER III

## SEQUENTIAL ESTIMATION OF $\mu$ FOR THE ONE-PARAMETER NEGATIVE BINOMIAL DISTRIBUTION

As mentioned in Chapter II, when $\mu$ is the single unknown parameter, the negative binomial family of distributions belongs to the exponential family and has a complete sufficient statistic, $\sum_{i=1}^{n} x_i$. Define

$$V_n = \frac{nk}{nk+1}\overline{X}_n + \frac{n}{nk+1}\overline{X}_n^2 \quad .$$

Since $E(V_n) = \sigma^2 = \mu + \frac{\mu^2}{k}$ and since $V_n$ is a function of the complete sufficient statistic, $V_n$ is the minimum variance unbiased estimator of $\sigma^2$. Using $V_n$ and $\overline{X}_n$ as the estimators of $\sigma^2$ and $\mu$, respectively, we shall develop three sequential procedures for estimating the mean in this chapter.

### Controlling the Coefficient of Variation of $\overline{X}$

After n observations, $x_1, x_2, \ldots, x_n$, suppose the loss incurred by estimating $\mu$ by $\overline{X}_n$ is given by

$$L_n = \frac{(\overline{X}_n - \mu)^2}{\mu^2} \quad .$$

The associated risk is then

$$\begin{aligned} R_n &= E(L_n) = \frac{\sigma^2/n}{\mu^2} \\ &= C^2 \end{aligned}$$

where C is the coefficient of variation of the mean. If $C^2$ (or equivalently C) is specified, then the fixed sample size required to achieve the desired risk is

$$n^*(C) = \left(\frac{\sigma}{\mu C}\right)^2 \quad . \tag{3.1}$$

Since $\mu$ and $\sigma$ are unknown, no fixed sample size procedure will achieve the desired risk for all $\mu$, $\sigma$.

Consider a sample of random size N with risk

$$E(L_N) = \frac{E(\overline{X}_N - \mu)^2}{\mu^2} \quad .$$

Using (3.1) as a guide and substituting our estimates of $\sigma^2$ and $\mu$, we would work with samples of random size

$$N \geq \frac{V_N}{(\overline{X}_N C)^2} \quad . \tag{3.2}$$

In order for the right-hand side of the above equation to be well-defined and positive almost surely, we will require sampling to continue until at least one positive value has been observed. Simplifying (3.2) algebraically, we obtain

$$\begin{aligned} N &\geq \frac{1}{C^2}\left(\frac{Nk}{(Nk + 1)\overline{X}_N} + \frac{N}{Nk + 1}\right) \\ &> \frac{N}{C^2(Nk + 1)} \quad . \end{aligned} \tag{3.3}$$

Solving the inequality for N, we find

$$N > \frac{1}{C^2 k} - \frac{1}{k} \tag{3.4}$$

leading us to a minimum sample size. Therefore, we propose to sample sequentially until

$$N = \min\left(n \geq n_0 = \max\left(2, \left[\frac{1}{C^2 k} - \frac{1}{k}\right]\right) : n \geq \frac{V_n}{(C\overline{X}_n)^2} \text{ and at least one nonzero value has been observed}\right) \tag{3.5}$$

where [y] is the largest integer less than or equal to y. Taking advantage of the nature of $V_n$, we can rewrite (3.5) as stop when

$$n \geq n_0 \quad \text{and} \quad T_n \geq \frac{nk}{C^2(nk + 1) - 1} .$$

We shall now investigate the properties of this sequential process by presenting some lemmas and a theorem. The main contribution of the first lemma is that it assures us that the stopping criterion will be met with a finite sample size for any fixed, positive C. The fact that the ratio of the random sample size to the optimal fixed sample size tends to 1 almost surely as C approaches 0 is proven in Lemma 2, and Lemma 3 states that the procedure is asymptotically efficient.

<u>Lemma 1</u>: N is nonincreasing in C and $P(N < \infty) = 1$ for any fixed $C > 0$.

<u>Proof</u>: Let $N(C) = \dfrac{V_N}{(\bar{X}_N C)^2}$ . Then $N'(C) = \dfrac{-2V_N}{\bar{X}_N^2 C^3} \leq 0$ . Hence N is nonincreasing in C. Also

$$\begin{aligned} P(N < \infty) &= 1 - P(N = \infty) \\ &= 1 - \lim_{n\to\infty} P(N > n) \\ &= 1 - \lim_{n\to\infty} P\left(n \leq \frac{V_n}{(C\bar{X}_n)^2}\right) \\ &= 1 - P\left(\lim_{n\to\infty} n \leq \left(\frac{\sigma}{C\mu}\right)^2\right) \\ &= 1 . \end{aligned}$$

Note that $\lim_{C\to 0} N = \infty$ a.s. since $n_0 \to \infty$ as $C \to 0$ .

<u>Lemma 2</u>: $\lim_{C\to 0} \dfrac{N}{n^*} = 1$ a.s.

<u>Proof</u>:

<u>Case 1</u>: Suppose $\left[\dfrac{1}{C^2 k} - \dfrac{1}{k}\right] \leq 2$ . In this case, $n_0 = 2$. So we must have

$$N - 1 < \frac{V_{N-1}}{(C\bar{X}_{N-1})^2} + (2 - 1) .$$

This implies

$$N < \frac{V_{N-1}}{(C\bar{X}_{N-1})^2} + 2 \ .$$

Case 2: Suppose $\left[\frac{1}{C^2k} - \frac{1}{k}\right] > 2$ . In this case, $n_0 = \left[\frac{1}{C^2k} - \frac{1}{k}\right]$.
Note: $P(N = n_0) = 0$. The proof of this will be by contradiction.
Assume $P(N = n_0) > 0$. This implies it is possible to obtain

$$n_0 \geq \frac{V_N}{(C\bar{X}_N)^2}$$

$$= \frac{1}{C^2}\left(\frac{Nk}{Nk + 1}\frac{1}{\bar{X}_N} + \frac{N}{Nk + 1}\right)$$

$$> \frac{1}{C^2}\left(\frac{N}{Nk + 1}\right) \text{ where } N = n_0 \ . \tag{3.6}$$

Substituting $n_0$ for $N$ in (3.6) and rewriting the inequality, we have

$$n_0 > \frac{1}{C^2k} - \frac{1}{k} \ . \tag{3.7}$$

However,

$$n_0 = \left[\frac{1}{C^2k} - \frac{1}{k}\right] \leq \frac{1}{C^2k} - \frac{1}{k} \ . \tag{3.8}$$

So (3.7) and (3.8) give a contradiction which implies $P(N = n_0) = 0$.
Hence

$$N - 1 < \frac{V_{N-1}}{(C\bar{X}_{N-1})^2} \quad \text{since} \quad N \geq n_0 + 1 \ .$$

Combining Cases (1) and (2), we have

$$\frac{V_N}{(C\bar{X}_N)^2} \leq N < \frac{V_{N-1}}{(C\bar{X}_{N-1})^2} + 2 \ .$$

Dividing by $n^* = \left(\frac{\sigma}{\mu C}\right)^2$, we obtain

$$V_N\left(\frac{\mu}{\sigma\overline{X}_N}\right)^2 \leq \frac{N}{n^*} < V_{N-1}\left(\frac{\mu}{\sigma\overline{X}_{N-1}}\right)^2 + 2\left(\frac{C\mu}{\sigma}\right)^2 .$$

Using the fact $\lim_{C\to 0} N = \infty$ and invoking the Strong Law of Large Numbers, we find

$$\lim_{C\to 0} \frac{N}{n^*} = 1 .$$

<u>Lemma 3</u>: $\lim_{C\to 0} E\left(\frac{N}{n^*}\right) = 1$ .

<u>Proof</u>: Using Fatou's lemma and Lemma 2, we have

$$\begin{aligned} \liminf_{C\to 0} E\left(\frac{N}{n^*}\right) &\geq E\left(\liminf_{C\to 0} \frac{N}{n^*}\right) \\ &= 1 . \end{aligned} \tag{3.9}$$

We shall complete the proof using exponential bounds, a technique first presented by Mukhopadhyay (27). Let $\varepsilon > 0$ be given. Define

$$\beta = (1 + \varepsilon)n^* = (1 + \varepsilon)\left(\frac{\sigma}{\mu C}\right)^2 = (1 + \varepsilon)\ \frac{(k + \mu)}{\mu k C^2} .$$

Then $E(N) = \sum_{n=n_0}^{\infty} n\, P(N = n)$

$$\leq \sum_{n=n_0}^{\beta} (\beta + 1)\ P(N = n) + \sum_{n\geq\beta+1}^{\infty} n\ P(N = n)$$

$$\leq (\beta + 1)\ P(N \leq \beta + 1) + T(\beta)$$

where $T(\beta) = \sum_{n\geq\beta+1}^{\infty} n\ P(N = n)$. Thus

$$E\left(\frac{N}{n^*}\right) \leq \left(\frac{\beta + 1}{n^*}\right)P(N \leq \beta + 1) + \frac{T(\beta)}{n^*} . \tag{3.10}$$

Then for sufficiently small C, if $T(\beta) \leq L$ where L is a constant independent of $\beta$, Lemma 2 together with (3.10) would imply

$$\limsup_{C\to 0} E\left(\frac{N}{n^*}\right) \leq 1 + \varepsilon$$

which together with (3.9) gives the desired result.

Note from (3.5) that

$$\{N = n\} \subset \left\{ n - 1 < \frac{V_{n-1}}{(C\bar{X}_{n-1})^2} \right\} . \tag{3.11}$$

Now $n - 1 < \dfrac{V_{n-1}}{(C\bar{X}_{n-1})^2}$ implies that $C^2(n - 1) < \dfrac{1}{\bar{X}_{n-1}} + \dfrac{1}{k}$ , and this may be rewritten as

$$\sum_{i=1}^{n-1} x_i < \frac{k(n - 1)}{C^2k(n - 1) - 1} . \tag{3.12}$$

Let $q(n, c) = \dfrac{kn}{C^2kn - 1}$ .

Then $n \geq \beta = (1 + \varepsilon)\left(\dfrac{\sigma}{\mu C}\right)^2$ implies that $C^2kn - 1 \geq (1 + \varepsilon)\left(\dfrac{\sigma}{\mu}\right)^2 k - 1$.

Hence

$$\frac{kn}{C^2kn - 1} \leq \frac{kn}{(1 + \varepsilon)\left(\frac{\sigma}{\mu}\right)^2 k - 1} . \tag{3.13}$$

Define $a(n) = \dfrac{kn}{(1+ \varepsilon)\left(\frac{\sigma}{\mu}\right)^2 k - 1}$

$$= \frac{nk\mu}{\varepsilon(\mu + k) + k} .$$

Thus from (3.11), (3.12) and (3.13), we have

$$T(\beta) = \sum_{n\geq\beta+1}^{\infty} n\, P(N = n)$$

$$= \sum_{n\geq\beta}^{\infty} (n + 1)\, P(N = n + 1)$$

$$\leq \sum_{n\geq\beta}^{\infty} (n + 1)\, P\left(n < V_n\left(\frac{1}{C\bar{X}_n}\right)^2\right)$$

$$\leq \sum_{n\geq\beta}^{\infty} (n + 1)\, P\left(\sum_{i=1}^{n} x_i < \frac{nk}{(1 + \varepsilon)\left(\frac{\sigma}{\mu}\right)^2 k - 1}\right)$$

$$\leq \sum_{n \geq \beta}^{\infty} (n+1)\, P\left(t \sum_{i=1}^{n} x_i > t\, a(n)\right) \quad \text{for any } t < 0$$

$$\leq \sum_{n \geq \beta} (n+1)\, P\left(e^{-t\,a(n)}\, e^{t \sum_{i=1}^{n} x_i} > 1\right).$$

Therefore, by Chebyshev's inequality, we obtain for any $t < 0$

$$T(\beta) \leq \sum_{n \geq \beta}^{\infty} (n+1)\, e^{-t\,a(n)}\, E\left(e^{t \sum_{i=1}^{n} x_i}\right). \tag{3.14}$$

Since $x_i \sim NB(\mu, k)$, $\sum_{i=1}^{n} x_i \sim NB(n\mu, nk)$ and

$$E\left(e^{t \sum_{i=1}^{n} x_i}\right) = \left(1 + \frac{\mu(1 - e^t)}{k}\right)^{-nk}, \quad t < \ell n\, \frac{k + \mu}{\mu}\ .$$

Hence (3.14) becomes

$$T(\beta) \leq \sum_{n > \beta}^{\infty} (n+1)\, e^{-t\,a(n)} \left(1 + \frac{\mu(1 - e^t)}{k}\right)^{-nk} \quad \text{for any } t < 0\ .$$

So

$$T(\beta) \leq \sum_{n > \beta}^{\infty} (n+1) \inf_{t<0} \left(e^{-t\,a(n)} \left(1 + \frac{\mu(1 - e^t)}{k}\right)^{-nk}\right). \tag{3.15}$$

Taking the first and second derivatives, we find that

$$t = \ell n\, \frac{a(n)(k + \mu)}{\mu\big(nk + a(n)\big)}$$

is a minimum, and it can be verified that $t$ is less than 0. Hence (3.15) becomes

$$T(\beta) \leq \sum_{n > \beta}^{\infty} (n+1) \left[(1 + \varepsilon)^{\frac{(\mu+k)(\varepsilon+1)}{\varepsilon(\mu+k)+k}} \left(\frac{k}{\varepsilon(\mu + k) + k}\right)\right]^{nk}$$

$$= \sum_{n > \beta}^{\infty} b_n\ , \text{ say.}$$

Now $b_n^{1/n} \to \ell < 1$ since $(n+1)^{\frac{1}{n}} \to 1$

and

$$\left[(1 + \varepsilon)^{\frac{(\mu+k)(\varepsilon+1)}{\varepsilon(\mu+k)+k}} \left(\frac{k}{\varepsilon(\mu + k) + k}\right)\right]^k < 1 \text{ for all } \varepsilon > 0 . \qquad (3.16)$$

To verify (3.16), consider

$$f(\varepsilon) = (1 + \varepsilon)^{\frac{(\mu+k)(\varepsilon+1)}{\varepsilon(\mu+k)+k}} \left(\frac{k}{\varepsilon(\mu+k)+k}\right) .$$

Note that $f(0) = 1$. By differentiating $\ell n\ f$, we can show that $f$ is a strictly decreasing function of $\varepsilon$. These two facts give us that $f(\varepsilon) < 1$ for $\varepsilon > 0$.

Then $b_n^{1/n} \to \ell < 1$ implies that $\sum_{n\geq\beta}^{\infty} b_n \to L$, a constant independent of $\beta$. So

$$T(\beta) \leq \sum_{n\geq\beta}^{\infty} b_n \to L .$$

Using Lemma 2 and (3.10) with the above, we have

$$\limsup_{C\to 0} E\left(\frac{N}{n^*}\right) \leq 1 + \varepsilon .$$

This together with (3.9) gives the desired result.

The preceding lemmas will be used to prove the risk efficiency of the proposed procedure in the following theorem. The methods used in proving this theorem were developed by Mukhopadhyay (28).

<u>Theorem 1</u>: $\lim_{C\to 0} \frac{E(L_N)}{R_{n^*}(C)} = 1.$

Proof: Note that

$$\lim_{C\to 0} \frac{E(L_N)}{R_{n^*}(C)} = \frac{E(\overline{X}_N - \mu)^2}{(\mu C)^2} = \frac{E(\overline{X}_N - \mu)^2 n^*}{\sigma^2} . \qquad (3.17)$$

Observe that

$$(\bar{X}_N - \mu)^2 = \frac{\left(\sum_{i=1}^{N} x_i - N\mu\right)^2}{n^{*2}} + \frac{\left(\sum_{i=1}^{N} x_i - N\mu\right)^2}{n^{*2}} \left(\left(\frac{n^*}{N}\right)^2 - 1\right) .$$

Thus

$$\frac{(\bar{X}_N - \mu)^2 n^*}{\sigma^2} = \frac{\left(\sum_{i=1}^{N} x_i - N\mu\right)^2}{\sigma^2 n^*} + \frac{\left(\sum_{i=1}^{N} x_i - N\mu\right)^2}{\sigma^2 n^*} \left(\left(\frac{n^*}{N}\right)^2 - 1\right)$$

$$= \quad I \quad + \quad J \quad , \text{ say.}$$

Using a result due to Anscombe (4), we obtain

$$I \xrightarrow{\mathcal{L}} \{N(0, 1)\}^2 \quad \text{as } C \to 0 .$$

Then from a theorem due to Chow, Robbins, and Teicher (10), we have

$$E(I) = \frac{\sigma^2 E(N)}{\sigma^2 n^*} = \frac{E(N)}{n^*} .$$

Using Lemma 3, we have $E(I) \to 1$ as $C \to 0$. Hence the family $\{I\}$ is uniformly integrable in the positive parameter $C$ (23, p. 183).

Now from the stopping rule, we have

$$N \geq \frac{1}{C^2}\left(\frac{Nk}{Nk + 1} \frac{1}{\bar{X}_N} + \frac{N}{Nk + 1}\right) > \frac{1}{C^2}\left(\frac{1}{k + 1}\right) .$$

This implies

$$\frac{1}{N} < C^2 (k + 1) .$$

Also

$$n^* = \left(\frac{\sigma}{\mu C}\right)^2 = \frac{\mu + k}{\mu k C^2} .$$

These give us

$$\frac{n^*}{N} \leq \frac{(\mu + k)(k + 1)}{\mu k} .$$

Thus

$$-1 \leq \left(\frac{n^*}{N}\right)^2 - 1 \leq \left(\frac{n^*}{N}\right)^2 \leq \left[\frac{(\mu + k)(k + 1)}{\mu k}\right]^2 .$$

This, together with the fact that $\{I\}$ is uniformly integrable implies that $\{J\}$ is also uniformly integrable. Hence, using Lemma 2, we have $E(J) \to 0$ as $C \to 0$. Therefore, (3.17) becomes

$$\frac{E(\overline{X}_N - \mu)^2 n^*}{\sigma^2} = 1 .$$

We shall also present a theorem closely related to one by Starr and Woodroofe (35) but given here for completeness. Let $\{c_n\}$ be any sequence of constants, and let $n_0$ be any positive integer. Define a stopping time of the sequence $x_1, x_2, \ldots$ by

$$\begin{aligned} N &= \text{smallest integer } n \geq n_0 \text{ such that } \overline{X}_n \geq c_n \\ &= \infty \text{ if no such } n \text{ exists, i.e., if } \overline{X}_n < c_n \text{ for every } n_0 \leq n < \infty . \end{aligned}$$

We assume $P(N < \infty) = 1$ so that $\overline{X}_N$ is well-defined.

<u>Theorem 2</u>: If $E(\overline{X}_N)$ exists, then $E(\overline{X}_N) \geq E(X_1) = \mu$ .

<u>Proof</u>: Without loss of generality, assume $E(X_1) = 0$. For any $n \geq n_0$ and any $i = 1, 2, \ldots, n$, we therefore have

$$\begin{aligned} \int_{(N>n)} x_i \, dP &= \int \cdots \int_A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\alpha} x_i \, dF(x_i) dF(x_n) \cdots \\ &\qquad dF(x_{i+1}) \, dF(x_{i-1}) \cdots dF(x_1) \\ &\leq 0 \end{aligned} \tag{3.18}$$

where A denotes the set of values of $x_1, x_2, \ldots, x_{i-1}$, for which $N > i - 1$ and $\alpha = \min\left\{kc_k - \sum_{i=1,j=1}^{k} x_j\right\}$ .

It follows that for any $n \geq n_0$

$$\begin{aligned} \int_{(N>n)} \overline{X}_n \, dP &= \int_{(N>n)} x_{max} \, dP \\ &\leq 0 \end{aligned}$$

since $x_{max} = x_i$ for some $1 \leq i \leq n$ and from (3.18). Also since $X_n$ is independent of the event $N > n - 1$,

$$\int_{(N>n-1)} \overline{X}_n \, dP = \frac{n-1}{n} \int_{(N>n-1)} \overline{X}_{n-1} \, dP$$

$$\geq \int_{(N>n-1)} \overline{X}_{n-1} \, dP \qquad (n > n_0)$$

Thus for every $n \geq n_0$ ,

$$\int_{(N\leq n)} \overline{X}_N \, dP = \sum_{i=1}^{n-1} \int_{(N=i)} \overline{X}_i \, dP + \int_{(N>n-1)} \overline{X}_n \, dP - \int_{(N>n)} \overline{X}_n \, dP$$

$$\geq \sum_{i=1}^{n-1} \int_{(N=i)} \overline{X}_i \, dP + \int_{(N>n-1)} \overline{X}_n \, dP$$

$$\geq \sum_{i=n_0}^{n-1} \int_{(N=i)} \overline{X}_i \, dP + \int_{(N>n-1)} \overline{X}_{n-1} \, dP$$

$$= \sum_{i=n_0}^{n-2} \int_{(N=i)} \overline{X}_i \, dP + \int_{(N>n-2)} \overline{X}_{n-1} \, dP$$

$$\geq \cdots \geq \int_{(N=n_0)} \overline{X}_{n_0} \, dP + \int_{(N>n_0)} \overline{X}_{n_0+1} \, dP$$

$$\geq \int_{(N=n_0)} \overline{X}_{n_0} \, dP + \int_{(N>n_0)} \overline{X}_{n_0} \, dP$$

$$\geq \int \overline{X}_{n_0} \, dP$$

$$= E(\overline{X}_{n_0}) = 0 \, .$$

Hence,

$$E(\overline{X}_N) = \lim_{n\to\infty} \int_{(N\leq n)} \overline{X}_N \, dP$$

$$\geq \lim_{n\to\infty} E(\overline{X}_{n_0})$$

$$= 0 \, .$$

The lemmas and the first theorem have a drawback when considering applications. They are limit results and give us no idea of the behavior of the sampling procedure for moderate values of C. We used Monte Carlo methods to evaluate the process for moderate C and various combinations of $\mu$ and k. Random numbers, $y_1$, $y_2$, ..., were drawn from a uniform (0, 1) population, and $F^{-1}(y_i)$ was determined for each, where F is the cumulative distribution function of a NB($\mu$, k) random variable. Hence we obtained observations from the appropriate negative binomial distributions. This method was used in all of the simulations in this thesis.

The values of $\mu$ and k were each allowed to vary from one to five by increments of one since this is the range most commonly found in entomology. Five hundred samples were taken for each combination of the parameters. Observations were added to each sample until the stopping criterion (3.5) was met.

In Tables I and II, we have presented the results of the simulations for C = .3 and C = .1 . For each combination of the parameters, the optimal fixed sample size ($n^*$), the average random sample size ($\overline{N}$), and the estimated standard deviation of $\overline{N}$ ($s_{\overline{N}}$) are presented. Also, the average of the estimates of the mean, $\overline{\overline{X}}_N$ , and their estimated standard deviation, $s_{\overline{X}_N}$ , were used to calculate the estimate of the true $CV(\overline{X}_N)$, $\hat{CV}(\overline{X}_N) = \dfrac{s_{\overline{X}_N}}{\overline{\overline{X}}_N}$ .

It is interesting to note that the optimal fixed sample size is a symmetric function of $\mu$ and k. This follows since

$$n^*(C) = \left(\frac{\sigma}{\mu C}\right)^2$$

$$= \frac{1}{C^2}\left(\frac{1}{\mu} + \frac{1}{k}\right).$$

Thus for C = .3, $n^* = 13.33$ when $\mu = 1$ and $k = 5$ and when $\mu = 5$ and $k = 1$ .

In the simulations, $n^*$ and $\overline{N}$ are close even for a C as moderate as .3 . The estimated standard deviation of $\overline{N}$ decreases as the mean in-

TABLE I

SEQUENTIAL ESTIMATION OF THE MEAN DESIGNED TO OBTAIN $CV(\overline{X}_N) = .3$*

| $\mu$ | k | n* | $\overline{N}$ | $s_{\overline{N}}$ | $\overline{\overline{X}}_N$ | $s_{\overline{X}_N}$ | $\widehat{CV}(\overline{X}_N)$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 22.22 | 22.74 | .17 | 1.03 | .32 | .31 |
| | 2 | 16.67 | 17.25 | .16 | 1.08 | .33 | .31 |
| | 3 | 14.82 | 15.72 | .17 | 1.07 | .32 | .30 |
| | 4 | 13.89 | 14.94 | .16 | 1.05 | .32 | .31 |
| | 5 | 13.33 | 14.21 | .15 | 1.07 | .33 | .31 |
| 2 | 1 | 16.67 | 16.62 | .09 | 2.08 | .61 | .29 |
| | 2 | 11.11 | 11.54 | .08 | 2.05 | .56 | .27 |
| | 3 | 9.26 | 9.92 | .09 | 2.07 | .61 | .29 |
| | 4 | 8.33 | 8.94 | .08 | 2.10 | .61 | .29 |
| | 5 | 7.78 | 8.48 | .08 | 2.07 | .65 | .31 |
| 3 | 1 | 14.82 | 14.59 | .06 | 3.15 | .95 | .30 |
| | 2 | 9.26 | 9.70 | .06 | 3.04 | .87 | .28 |
| | 3 | 7.41 | 7.80 | .06 | 3.15 | .93 | .30 |
| | 4 | 6.48 | 6.99 | .06 | 3.17 | .87 | .28 |
| | 5 | 5.93 | 6.45 | .05 | 3.16 | .96 | .30 |
| 4 | 1 | 13.89 | 13.70 | .05 | 4.00 | 1.11 | .28 |
| | 2 | 8.33 | 8.59 | .04 | 4.11 | 1.15 | .28 |
| | 3 | 6.48 | 6.91 | .04 | 4.07 | 1.16 | .29 |
| | 4 | 5.56 | 6.06 | .04 | 4.10 | 1.17 | .29 |
| | 5 | 5.00 | 5.52 | .04 | 4.08 | 1.15 | .28 |
| 5 | 1 | 13.33 | 13.03 | .04 | 5.02 | 1.37 | .27 |
| | 2 | 7.78 | 8.01 | .04 | 5.06 | 1.46 | .29 |
| | 3 | 5.93 | 6.22 | .03 | 5.21 | 1.44 | .28 |
| | 4 | 5.00 | 5.47 | .04 | 5.13 | 1.52 | .30 |
| | 5 | 4.44 | 4.92 | .04 | 5.18 | 1.50 | .29 |

* Each entry is based on 500 simulations.

TABLE II

SEQUENTIAL ESTIMATION OF THE MEAN DESIGNED TO OBTAIN $CV(\bar{X}_N) = .1$*

| $\mu$ | k | n* | $\bar{N}$ | $s_{\bar{N}}$ | $\bar{\bar{X}}_N$ | $s_{\bar{X}_N}$ | $\widehat{CV}(\bar{X}_N)$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 200.00 | 201.13 | .47 | 1.00 | .10 | .10 |
| | 2 | 150.00 | 150.67 | .45 | 1.01 | .10 | .10 |
| | 3 | 133.33 | 134.21 | .44 | 1.01 | .10 | .10 |
| | 4 | 125.00 | 126.56 | .46 | 1.00 | .10 | .10 |
| | 5 | 120.00 | 121.44 | .45 | 1.00 | .10 | .10 |
| 2 | 1 | 150.00 | 150.01 | .24 | 2.00 | .21 | .10 |
| | 2 | 100.00 | 100.70 | .23 | 2.00 | .21 | .10 |
| | 3 | 83.33 | 83.85 | .22 | 2.01 | .20 | .10 |
| | 4 | 75.00 | 75.75 | .22 | 2.01 | .19 | .10 |
| | 5 | 70.00 | 70.87 | .23 | 2.00 | .20 | .10 |
| 3 | 1 | 133.33 | 133.09 | .16 | 3.01 | .31 | .10 |
| | 2 | 83.33 | 83.96 | .15 | 2.98 | .30 | .10 |
| | 3 | 66.67 | 67.04 | .17 | 3.02 | .33 | .11 |
| | 4 | 58.33 | 58.96 | .16 | 3.01 | .31 | .10 |
| | 5 | 53.33 | 54.02 | .15 | 3.00 | .30 | .10 |
| 4 | 1 | 125.00 | 124.93 | .11 | 3.98 | .39 | .10 |
| | 2 | 75.00 | 75.19 | .11 | 4.01 | .40 | .10 |
| | 3 | 58.33 | 58.82 | .12 | 3.99 | .41 | .10 |
| | 4 | 50.00 | 50.55 | .12 | 4.01 | .41 | .10 |
| | 5 | 45.00 | 45.68 | .12 | 3.99 | .39 | .10 |
| 5 | 1 | 120.00 | 119.79 | .09 | 4.99 | .49 | .10 |
| | 2 | 70.00 | 70.04 | .09 | 5.05 | .50 | .10 |
| | 3 | 53.33 | 53.68 | .09 | 5.01 | .51 | .10 |
| | 4 | 45.00 | 45.40 | .09 | 5.02 | .51 | .10 |
| | 5 | 40.00 | 40.37 | .09 | 5.05 | .51 | .10 |

* Each entry is the result of 500 simulations.

creases, but it is less than one-half for every case considered.

The possible positive bias stated in Theorem 2 is small but noticeable when C = .3, but appears to be negligible for C = .1 . Lastly, the $CV(\overline{X}_N)$ is close to the stated level even for C = .3 . There is a tendency for the estimated $CV(\overline{X}_N)$ to be slightly higher than the stated C = .3 when the mean is small. However, when rounded to two decimal places, $\widehat{CV}(\overline{X}_N)$ is equal to the specified C = .1 for all but one of the combinations of the parameters considered.

There may be times when we are sampling from the negative binomial distribution and believe we know k, but our knowledge of k is imprecise. Considering the stopping rule as a function of k, we have

$$f(k) = \frac{Nk}{C^2(Nk + 1) - 1} .$$

Taking the first derivative, we have $f'(k) < 0$ for all $C < 1$. So as k increases, we may stop with smaller values of $T_n$ . Hence, to be conservative, we would want to underestimate k.

If our value of k is not exact, how much does that affect our estimates? This question was studied some, and the results are in Table III. The true value of k is 2 and the mean is 1, but we used stopping rules based on k from 1.1 to 3 by increments of .1 . We specified C = .1, and thus the optimal fixed sample size is 150. Notice that we do tend to take more observations when k is less than 2 and fewer when k is greater than 2. However, the value of $\widehat{CV}(\overline{X}_N)$ is not affected greatly. Thus slight misses in the value of k do not seem to invalidate the sampling process.

Inspection of the proofs will show that all of the lemmas and theorems proven hold when Kuno's procedure is applied to the negative

TABLE III

STUDY OF THE EFFECT OF IMPRECISE KNOWLEDGE OF k WHEN THE GOAL IS TO CONTROL $CV(\bar{X}_N) = .1$*

| $\mu$ | True Value of k | Assumed Value of k | $\bar{N}$ | $s_{\bar{N}}$ | $\bar{\bar{X}}_N$ | $s_{\bar{X}_N}$ | $\hat{CV}(\bar{X}_N)$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 1.1 | 191.74 | .42 | 1.00 | .09 | .09 |
| | | 1.2 | 183.69 | .41 | 1.00 | .09 | .09 |
| | | 1.3 | 177.33 | .40 | 1.00 | .09 | .09 |
| | | 1.4 | 172.47 | .43 | 1.00 | .09 | .09 |
| | | 1.5 | 167.57 | .43 | 1.00 | .10 | .10 |
| 1 | 2 | 1.6 | 163.10 | .44 | 1.01 | .10 | .10 |
| | | 1.7 | 160.18 | .47 | 1.00 | .10 | .10 |
| | | 1.8 | 155.88 | .45 | 1.01 | .10 | .10 |
| | | 1.9 | 153.88 | .44 | 1.00 | .10 | .10 |
| | | 2.0 | 150.33 | .45 | 1.01 | .10 | .10 |
| 1 | 2 | 2.1 | 149.19 | .47 | 1.00 | .10 | .10 |
| | | 2.2 | 146.30 | .45 | 1.01 | .10 | .10 |
| | | 2.3 | 144.43 | .45 | 1.00 | .10 | .10 |
| | | 2.4 | 142.87 | .47 | 1.00 | .10 | .10 |
| | | 2.5 | 140.99 | .47 | 1.01 | .10 | .10 |
| 1 | 2 | 2.6 | 139.68 | .47 | 1.00 | .10 | .10 |
| | | 2.7 | 138.58 | .50 | 1.00 | .11 | .11 |
| | | 2.8 | 136.19 | .47 | 1.01 | .11 | .11 |
| | | 2.9 | 135.81 | .47 | 1.00 | .11 | .11 |
| | | 3.0 | 133.81 | .47 | 1.01 | .11 | .10 |

* Each entry is based on 500 simulations.

binomial distribution. Lemmas 1 and 2 and Theorem 2 are also valid when considering any distribution where the variance is a quadratic function of the mean. However, the proof of Lemma 3, and consequently the one for Theorem 1, involve the moment generating function of the negative binomial, and we have been unable thus far to extend it to the more general case stated by Kuno.

Upon examination of the stopping rules, we can see that our procedure may require slightly fewer observations than Kuno's. In Table IV, we have presented the results of simulations based on Kuno's stopping rule for the negative binomial. Each entry is the result of 500 simulations, and the desired C is .3 . We do note that the average sample sizes are consistently higher, and as a result, the $\hat{CV}(\bar{X}_N)$ tends to be smaller than with our procedure.

A few attempts were made to simulate the methods used by Allen, Gonzalez, and Gokhale (1). We considered the variance as a quadratic function of the mean, $a\mu + b\mu^2$. Drawing samples from the negative binomial, we estimated a and b by regression, and then employed Kuno's procedure. The estimates of a and b generally proved to be very poor, and consequently, the goal of obtaining a desired C was missed.

## Estimation of $\mu$ Within $p\mu$ with Confidence $1 - \alpha$

We shall again be considering the one-parameter negative binomial distribution. Nadás (29) speaks of proportional accuracy when estimating $\mu$ by

$$J_n = (\mu: |\bar{X}_n - \mu| \leq p|\mu|) . \tag{3.19}$$

If for a given p, we want $J_n$ to cover $\mu$ with probability $1 - \alpha$, then upon invoking the Central Limit Theorem, the required fixed sample size is

TABLE IV

KUNO'S SEQUENTIAL ESTIMATION OF THE MEAN DESIGNED TO OBTAIN CV($\overline{X}_N$) = .3*

| $\mu$ | k | n* | $\overline{N}$ | $s_{\overline{N}}$ | $\overline{\overline{X}}_N$ | $s_{\overline{X}_N}$ | $\widehat{CV}(\overline{X}_N)$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 22.22 | 23.69 | .16 | 1.02 | .29 | .29 |
| | 2 | 16.67 | 17.82 | .16 | 1.06 | .32 | .30 |
| | 3 | 14.82 | 15.92 | .15 | 1.06 | .31 | .29 |
| | 4 | 13.89 | 14.97 | .16 | 1.07 | .32 | .30 |
| | 5 | 13.33 | 14.24 | .15 | 1.09 | .33 | .31 |
| 2 | 1 | 16.67 | 17.50 | .08 | 2.07 | .58 | .28 |
| | 2 | 11.11 | 11.92 | .08 | 2.12 | .61 | .29 |
| | 3 | 9.26 | 9.98 | .08 | 2.12 | .60 | .28 |
| | 4 | 8.33 | 9.39 | .08 | 2.05 | .60 | .29 |
| | 5 | 7.78 | 8.61 | .08 | 2.10 | .62 | .29 |
| 3 | 1 | 14.82 | 15.75 | .06 | 2.95 | .85 | .29 |
| | 2 | 9.26 | 10.01 | .06 | 3.11 | .89 | .29 |
| | 3 | 7.41 | 8.11 | .05 | 3.13 | .90 | .29 |
| | 4 | 6.48 | 7.35 | .06 | 3.07 | .98 | .32 |
| | 5 | 5.93 | 6.68 | .05 | 3.10 | .87 | .28 |
| 4 | 1 | 13.89 | 14.58 | .04 | 4.08 | 1.19 | .29 |
| | 2 | 8.33 | 9.04 | .04 | 4.13 | 1.21 | .29 |
| | 3 | 6.48 | 7.18 | .04 | 4.16 | 1.23 | .30 |
| | 4 | 5.56 | 6.24 | .04 | 4.10 | 1.09 | .27 |
| | 5 | 5.00 | 5.64 | .04 | 4.18 | 1.14 | .27 |
| 5 | 1 | 13.33 | 14.01 | .04 | 5.07 | 1.45 | .29 |
| | 2 | 7.78 | 8.49 | .03 | 5.05 | 1.43 | .28 |
| | 3 | 5.93 | 6.60 | .03 | 5.14 | 1.54 | .30 |
| | 4 | 5.00 | 5.66 | .03 | 5.07 | 1.46 | .29 |
| | 5 | 4.44 | 5.04 | .03 | 5.28 | 1.47 | .28 |

* Each entry is based on 500 simulations.

$$n^*(p) = \left(\frac{z\sigma}{p\mu}\right)^2 .$$

If $\sigma$ or $\mu$ is unknown, then we cannot determine $n^*$. Again using $V_n$ and $\bar{X}_n$ to estimate $\sigma$ and $\mu$, respectively, we shall consider the stopping rule

$$N = \min\left(n \geq n_0 = \max\left(2, \left[\frac{z^2}{p^2 k} - \frac{1}{k}\right]\right) : n \geq V_n\left(\frac{z}{\bar{X}_n p}\right)^2 \text{ and at least one nonzero value has been observed}\right) . \tag{3.20}$$

The minimum sample size requirement follows from algebra similar to that demonstrated in (3.3) and (3.4). Sampling is not allowed to stop before observing one positive quantity, assuring us that $V_N\left(\frac{z}{\bar{X}_N p}\right)^2$ is well-defined and positive almost surely. The stopping time N is well-defined, and we can rewrite (3.20) as stop when

$$n \geq n_0 \quad \text{and} \quad T_n \geq \frac{nkz^2}{(nk + 1)p^2 - z^2} .$$

The following lemma indicates N tends to increase as p becomes smaller, and the stopping criterion will be met with a finite sample size for any positive p. Since the proof has only minor differences from that of Lemma 1, we shall not include it here.

Lemma 4: N is nonincreasing in p and $P(N < \infty) = 1$ for any fixed $p > 0$ . Note that $\lim_{p\to 0} N = \infty$ a.s. since $n_0 \to \infty$ as $p \to 0$.

The properties of this sequential process will be further explored in the following theorem. Since the proofs of (3.21) and (3.23) closely parallel those of Lemmas 2 and 3, they will be omitted.

Theorem 3: Consider the interval estimate of $\mu$ by $J_n$ in (3.19). Then

$$\lim_{p\to 0} \frac{N}{n^*} = 1 \quad \text{a.s.} \tag{3.21}$$

$$\lim_{p\to 0} P(\mu \in J_n) = 1 - \alpha \quad \text{"asymptotic consistency"} \tag{3.22}$$

$$\lim_{p\to 0} E\left(\frac{N}{n^*}\right) = 1 \quad \text{"asymptotic efficiency"} \tag{3.23}$$

Proof of (3.22): Since $n^*(p) = \left(\frac{z\sigma}{p\mu}\right)^2$, we can rewrite (3.21) as $\lim_{p\to 0} \frac{Np\mu}{\sigma} = z$. Now

$$P(\mu \in J_n) = P(|\bar{X}_n - \mu| \leq p|\mu|)$$

$$= P\left(\frac{|X_1 + X_2 + \ldots + X_N - N\mu|}{\sigma\sqrt{N}} \leq \frac{p\mu\sqrt{N}}{\sigma}\right) .$$

Since $\frac{p\mu\sqrt{N}}{\sigma} \to z$ and $\frac{N}{n^*} \to 1$ in probability as $p \to 0$, it then follows from a result of Anscombe (4) that as $p \to 0$,

$$\frac{X_1 + X_2 + \ldots X_N - N\mu}{\sigma\sqrt{N}} \sim N(0, 1) .$$

Hence

$$\lim_{p\to 0} P(\mu \in J_n) = \int_{-z}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$= 1 - \alpha$$

which proves (3.22).

We should note that Theorem 2 is also applicable to this sequential process.

This procedure was investigated, using simulation, to determine its behavior for moderate values of $p$. Tables V and VI present the results for $p = .3$ and $p = .2$ where the stated level of confidence is .95 .

As was the case when our goal was to attain a specified $CV(\bar{X})$, the optimal fixed sample size is a symmetric function of $\mu$ and $k$. Notice that $n^*$, the optimal fixed sample size, and $\bar{N}$, the average random sample size, are close. The estimated standard deviation of $\bar{N}$, $s_{\bar{N}}$, is less than one-half in every instance.

Although our goal is to estimate $\mu$ within $p\mu$, we believe that in

TABLE V

SEQUENTIAL ESTIMATION OF μ DESIGNED TO ESTIMATE μ WITHIN .3μ WITH 95% CONFIDENCE*

| $\mu$ | k | n* | $\overline{N}$ | $s_{\overline{N}}$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \le p\mu)$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \le p\overline{X}_N)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 85.37 | 86.17 | .30 | .954 | .952 |
| | 2 | 64.03 | 64.30 | .29 | .952 | .950 |
| | 3 | 56.19 | 57.45 | .30 | .934 | .958 |
| | 4 | 53.36 | 54.29 | .29 | .952 | .952 |
| | 5 | 51.22 | 51.99 | .29 | .928 | .944 |
| 2 | 1 | 64.03 | 64.20 | .16 | .942 | .926 |
| | 2 | 42.68 | 43.07 | .15 | .964 | .948 |
| | 3 | 35.57 | 35.82 | .14 | .944 | .954 |
| | 4 | 32.01 | 32.77 | .14 | .960 | .962 |
| | 5 | 29.88 | 30.59 | .16 | .936 | .948 |
| 3 | 1 | 56.91 | 56.44 | .10 | .930 | .952 |
| | 2 | 35.57 | 35.70 | .10 | .944 | .958 |
| | 3 | 28.46 | 28.83 | .11 | .932 | .932 |
| | 4 | 24.90 | 25.48 | .10 | .962 | .966 |
| | 5 | 22.76 | 23.33 | .11 | .934 | .948 |
| 4 | 1 | 53.36 | 53.07 | .08 | .950 | .950 |
| | 2 | 32.01 | 32.18 | .08 | .950 | .956 |
| | 3 | 24.90 | 25.21 | .07 | .952 | .960 |
| | 4 | 21.34 | 21.89 | .08 | .948 | .948 |
| | 5 | 19.21 | 19.72 | .07 | .954 | .956 |
| 5 | 1 | 51.22 | 50.90 | .06 | .968 | .950 |
| | 2 | 29.88 | 30.05 | .06 | .954 | .940 |
| | 3 | 22.76 | 23.18 | .06 | .968 | .938 |
| | 4 | 19.21 | 19.66 | .06 | .946 | .936 |
| | 5 | 17.07 | 17.57 | .06 | .952 | .964 |

* Each entry is the result of 500 simulations.

TABLE VI

SEQUENTIAL ESTIMATION OF μ DESIGNED TO ESTIMATE μ WITHIN .2μ WITH 95% CONFIDENCE*

| $\mu$ | k | n* | $\overline{N}$ | $s_{\overline{N}}$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \leq p\mu)$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \leq p\overline{X}_N)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 192.08 | 192.34 | .46 | .942 | .948 |
| | 2 | 144.06 | 144.61 | .44 | .950 | .950 |
| | 3 | 128.05 | 128.80 | .44 | .944 | .962 |
| | 4 | 120.05 | 120.67 | .43 | .944 | .952 |
| | 5 | 115.25 | 116.85 | .43 | .964 | .954 |
| 2 | 1 | 144.06 | 143.80 | .23 | .944 | .936 |
| | 2 | 96.04 | 96.51 | .23 | .942 | .944 |
| | 3 | 80.03 | 80.38 | .23 | .952 | .948 |
| | 4 | 72.03 | 72.68 | .21 | .954 | .964 |
| | 5 | 67.23 | 67.96 | .22 | .960 | .950 |
| 3 | 1 | 128.05 | 128.01 | .15 | .956 | .946 |
| | 2 | 80.03 | 80.50 | .15 | .952 | .940 |
| | 3 | 64.03 | 64.42 | .15 | .952 | .944 |
| | 4 | 56.02 | 56.67 | .15 | .956 | .954 |
| | 5 | 51.22 | 51.66 | .14 | .956 | .960 |
| 4 | 1 | 120.05 | 119.76 | .11 | .952 | .950 |
| | 2 | 72.03 | 72.17 | .11 | .956 | .950 |
| | 3 | 56.02 | 56.09 | .11 | .956 | .962 |
| | 4 | 48.02 | 48.45 | .12 | .934 | .944 |
| | 5 | 43.22 | 43.88 | .11 | .950 | .952 |
| 5 | 1 | 115.25 | 115.03 | .09 | .932 | .932 |
| | 2 | 67.23 | 67.32 | .09 | .956 | .950 |
| | 3 | 51.22 | 51.56 | .09 | .950 | .954 |
| | 4 | 43.22 | 43.72 | .09 | .932 | .926 |
| | 5 | 38.42 | 38.90 | .09 | .938 | .942 |

* Each entry is based on 500 simulations.

practice many will state the conclusions as having estimated $\mu$ within $p\overline{X}_N$. Thus we computed the observed level of confidence for both statements. There are some differences, but they are not consistent and may be due to variability in the estimates. Both vary about the desired level of .95.

We studied Binns' procedure using simulation. Again each entry is based on 500 simulations, and the results for $p = .2$ and $\alpha = .05$ are presented in Table VII. Slightly higher sample sizes are consistently required when using his method than when using ours, but the levels of confidence differ only in a random fashion.

The disadvantage in Binns' process is the interpolation required to obtain the final estimate, and the benefit is the asymptotic log normality of the estimate for large a. Using the univariate procedure in the Statistical Analysis System (S.A.S.), we considered the normality of the logs of the estimates for the two methods. We found that for $\mu = 1$, $k = 2$, $p = .2$ and $\alpha = .05$, we were unable to reject the null hypothesis of normality for either procedure after drawing 500 samples. The observed significance level was slightly smaller for our procedure, however.

## Estimation of $\mu$ Within d with Confidence $1 - \alpha$

Suppose now we want to estimate $\mu$ by

$$I_n = (\mu\colon |\overline{X}_n - \mu| \leq d) \tag{3.24}$$

when sampling from a one-parameter negative binomial distribution. If, for a given d, we want $I_n$ to cover $\mu$ with probability $1 - \alpha$, then after invoking the Central Limit Theorem, we determine the fixed sample size to be

TABLE VII

BINNS' SEQUENTIAL PROCEDURE DESIGNED TO ESTIMATE $\mu$ WITHIN $.2\mu$ WITH 95% CONFIDENCE *

| $\mu$ | k | n* | $\overline{N}$ | $s_{\overline{N}}$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \leq p\mu)$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \leq p\overline{X}_N)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 192.08 | 195.43 | .46 | .940 | .950 |
| | 2 | 144.06 | 146.60 | .44 | .950 | .948 |
| | 3 | 128.05 | 130.48 | .44 | .948 | .964 |
| | 4 | 120.05 | 122.19 | .43 | .952 | .950 |
| | 5 | 115.25 | 118.27 | .43 | .960 | .952 |
| 2 | 1 | 144.06 | 146.63 | .23 | .942 | .942 |
| | 2 | 96.04 | 98.04 | .23 | .948 | .946 |
| | 3 | 80.03 | 81.50 | .22 | .948 | .954 |
| | 4 | 72.03 | 73.73 | .21 | .958 | .958 |
| | 5 | 67.23 | 68.84 | .22 | .954 | .952 |
| 3 | 1 | 128.05 | 130.62 | .14 | .952 | .950 |
| | 2 | 80.03 | 81.92 | .15 | .950 | .948 |
| | 3 | 64.03 | 65.46 | .15 | .950 | .948 |
| | 4 | 56.02 | 57.53 | .15 | .954 | .954 |
| | 5 | 51.22 | 52.39 | .14 | .960 | .962 |
| 4 | 1 | 120.05 | 122.37 | .11 | .952 | .956 |
| | 2 | 72.03 | 73.55 | .11 | .956 | .956 |
| | 3 | 56.02 | 57.08 | .11 | .950 | .968 |
| | 4 | 48.02 | 49.20 | .12 | .938 | .944 |
| | 5 | 43.22 | 44.47 | .11 | .946 | .952 |
| 5 | 1 | 115.25 | 117.59 | .09 | .942 | .936 |
| | 2 | 67.23 | 68.68 | .09 | .956 | .956 |
| | 3 | 51.22 | 52.46 | .09 | .952 | .954 |
| | 4 | 43.22 | 44.47 | .09 | .938 | .926 |
| | 5 | 38.42 | 39.48 | .09 | .948 | .948 |

* Each entry is the result of 500 simulations.

$$n^*(d) = \left(\frac{z\sigma}{d}\right)^2 .$$

Since $\mu$, and therefore $\sigma$, is unknown, we cannot determine $n^*$. So consider the sequential procedure with a random sample size

$$N = \min\left(n \geq n_0 \ (\geq 2): n \geq V_n\left(\frac{z}{d}\right)^2 \text{ and at least one nonzero value has been observed}\right) . \tag{3.25}$$

The stopping time N is well-defined, and (3.25) may be restated as stop when

$$n \geq n_0 \quad \text{and} \quad n\sqrt{(nk + 1)\left(\frac{d}{z}\right)^2 + \frac{k^2}{4}} - \frac{nk}{2} \geq T_n .$$

Since the following lemma and theorems are proven, with only minor changes, as earlier ones, we shall omit the proofs.

Lemma 5: N is nonincreasing in d and $P(N < \infty) = 1$ for any fixed $d > 0$.

Theorem 4: Consider the interval estimate of $\mu$ by $I_n$ in (3.20). Then

$$\lim_{d\to 0} \frac{N}{n^*} = 1 \quad \text{a.s.} \tag{3.22}$$

$$\lim_{d\to 0} P(\mu \in I_N) = 1 - \alpha \qquad \text{"asymptotic consistency"} \tag{3.23}$$

$$\lim_{d\to 0} E\left(\frac{N}{n^*}\right) = 1 \qquad \text{"asymptotic efficiency"} \tag{3.24}$$

Theorem 5: Let $\{c_n\}$ be any sequence of constants, and let $n_0$ be any positive integer. Define a stopping time of the sequence $x_1, x_2, \ldots$ by

$N$ = smallest integer $n \geq n_0$ such that $\overline{X}_n \leq c_n$

$= \infty$ if no such n exists, i.e., if $\overline{X}_n > c_n$ for every $n_0 \leq n < \infty$.

We assume $P(N < \infty) = 1$ so that $\overline{X}_N$ is well-defined. If $E(\overline{X}_N)$ exists, $E(\overline{X}_N) \leq E(X_1) = \mu$. Theorem 5 is due to Starr and Woodroofe (35).

Simulation results for $d = .5$ and $\alpha = .05$ are in Table VIII. Notice that the estimated confidence level is far below the stated one

TABLE VIII

SEQUENTIAL ESTIMATION OF μ DESIGNED TO ESTIMATE μ WITHIN .5 UNITS WITH 95% CONFIDENCE *

| $\mu$ | k | n* | $\overline{N}$ | $s_{\overline{N}}$ | Estimated $P(\lvert\overline{X}_N - \mu\rvert \leq d)$ |
|---|---|---|---|---|---|
| 1 | 1 | 30.73 | 23.12 | .62 | .732 |
| | 2 | 23.05 | 19.38 | .41 | .818 |
| | 3 | 20.49 | 17.95 | .37 | .812 |
| | 4 | 19.21 | 17.15 | .31 | .832 |
| | 5 | 18.44 | 16.44 | .31 | .806 |
| 2 | 1 | 92.20 | 82.68 | 1.28 | .858 |
| | 2 | 61.47 | 57.41 | .67 | .916 |
| | 3 | 51.22 | 50.04 | .48 | .924 |
| | 4 | 46.10 | 45.20 | .38 | .948 |
| | 5 | 43.03 | 42.56 | .36 | .928 |
| 3 | 1 | 184.40 | 179.35 | 1.62 | .934 |
| | 2 | 115.25 | 112.87 | .75 | .926 |
| | 3 | 92.20 | 91.30 | .56 | .942 |
| | 4 | 80.67 | 80.50 | .45 | .946 |
| | 5 | 73.76 | 73.02 | .41 | .932 |
| 4 | 1 | 307.33 | 305.12 | 1.69 | .942 |
| | 2 | 184.40 | 183.61 | .92 | .954 |
| | 3 | 143.42 | 142.73 | .65 | .934 |
| | 4 | 122.93 | 122.57 | .55 | .936 |
| | 5 | 110.64 | 109.59 | .48 | .960 |
| 5 | 1 | 460.99 | 444.19 | 2.00 | .950 |
| | 2 | 268.91 | 268.76 | 1.00 | .964 |
| | 3 | 204.88 | 204.26 | .82 | .924 |
| | 4 | 172.87 | 172.38 | .71 | .948 |
| | 5 | 153.66 | 154.13 | .52 | .966 |

* Each entry is the result of 500 simulations.

when $\mu = 1$, and it tends to be below the stated level for other values of $\mu$ as well. This is due, at least in part, to the fact that $\bar{N}$ is consistently lower than $n^*$ .

In Table IX, we have the results when $\mu = 1$, $d = .2$, and $\alpha = .05$ . For smaller values of d, the observed confidence level is much closer to the stated one. $\bar{N}$ and $n^*$ are closer together in this case than they were when $d = .5$ .

TABLE IX

SEQUENTIAL ESTIMATION OF $\mu$ DESIGNED TO ESTIMATE $\mu$ WITHIN .2 UNITS WITH 95% CONFIDENCE*

| $\mu$ | k | $n^*$ | $\bar{N}$ | $s_{\bar{N}}$ | Estimated $P(\lvert\bar{X}_N - \mu\rvert \leq d)$ |
|---|---|---|---|---|---|
| 1 | 1 | 192.08 | 187.65 | 1.34 | .952 |
| | 2 | 144.06 | 142.62 | .91 | .940 |
| | 3 | 128.05 | 128.27 | .72 | .958 |
| | 4 | 120.05 | 119.61 | .69 | .932 |
| | 5 | 115.25 | 114.72 | .61 | .948 |

* Each entry is based on 500 simulations.

# CHAPTER IV

## NONPARAMETRIC, SEQUENTIAL ESTIMATION OF $\mu$ APPLIED TO THE NEGATIVE BINOMIAL DISTRIBUTION

Three different sequential procedures were studied in Chapter III. In the two cases where the goal is an interval estimate of $\mu$ with specified closeness and confidence, parallel procedures exist in the literature to cover the case where the distribution is unspecified. We have found no reference to a nonparametric process for estimating $\mu$ with prescribed coefficient of variation of the mean, although this is similar to the problem considered first by Mukhopadhyay (28). Such a procedure would be beneficial if there is some doubt as to the adequacy of the fit of the negative binomial to the population of interest or if k is unknown.

Following the same notation used in Chapter III, assume the loss incurred by estimating $\mu$ by $\overline{X}_n$ is given by

$$L_n = \frac{(\overline{X}_n - \mu)^2}{\mu^2} \quad .$$

The associated risk is then

$$\begin{aligned} R_n &= E(L_n) = \frac{\sigma^2/n}{\mu^2} \\ &= C^2 \end{aligned}$$

where C is the coefficient of variation of the mean. If C is specified in advance, the required fixed sample size is

$$n^*(C) = \left(\frac{\sigma}{\mu C}\right)^2 .$$

Since no knowledge of the distribution is assumed, $\mu$ and $\sigma$ are unknown, and thus $n^*$ cannot be determined.

First assume that $0 < \rho_0 \leq \rho < \infty$ for a known positive constant $\rho_0$ where $\rho^2 = \left(\frac{\sigma}{\mu}\right)^2$. It is not necessary for $\rho_0$ to be close to $\rho$, but a lower bound is required. If $\rho$ is unknown, then $\rho_0$ may be chosen to be arbitrarily small. Now consider a sample of random size N with risk

$$E(L_N) = \frac{E(\overline{X}_N - \mu)^2}{\mu^2} .$$

We propose to use $\overline{X}_N$ as the estimate of $\mu$ where

$$N = \min\left(n \geq n_0 = \max\left(2, \left\{\left(\frac{\rho_0}{C}\right)^2\right\}\right) : n \geq \frac{1}{C^2} \max\left(\left(\frac{u_n}{\overline{X}_n}\right)^2, \rho_0^2\right)\right). \quad (4.1)$$

Here $\{y\}$ denotes the smallest integer less than y. The stopping time N is well-defined.

The proofs of Lemmas 6 and 7 correspond closely to those of Lemmas 1 and 2 and will, therefore, be omitted.

Lemma 6: N is nonincreasing in C and $P(N < \infty) = 1$ for any fixed $C > 0$.

Lemma 7: $\lim_{C \to 0} \left(\frac{N}{n^*}\right) = 1$ a.s.

The asymptotic efficiency of the sequential procedure is considered in the next lemma. The second case of the proof differs only slightly from the one for proportional accuracy given by Nadás (29), but we have included it here for completeness.

Lemma 8: $\lim_{C \to 0} E\left(\frac{N}{n^*}\right) = 1$ .

Using Lemma 7 and Fatou's lemma, we have

$$\liminf_{C\to 0} E\left(\frac{N}{n^*}\right) \geq E\left(\liminf_{C\to 0} \frac{N}{n^*}\right) \tag{4.2}$$
$$= 1 .$$

To prove the lim sup part, we shall consider two cases.

Case 1: Suppose $\max\left(\rho_0^2, \left(\frac{u_n}{\bar{X}_n}\right)^2\right) = \rho_0^2$. Then from (4.1) we have

$$0 < N < \left(\frac{\rho_0}{C}\right)^2 + 2 . \tag{4.3}$$

Dividing through by $n^* = \left(\frac{\sigma}{\mu C}\right)^2$, (4.3) becomes

$$0 < \frac{N}{n^*} < \left(\frac{\mu}{\sigma}\right)^2 (\rho_0^2 + 2) .$$

Since $\frac{N}{n^*}$ is dominated by an integrable function, we have

$$\limsup_{C\to 0} E\left(\frac{N}{n^*}\right) \leq 1$$

by Lemma 2 and the Lebesque Dominated Convergence Theorem.

Case 2: Suppose $\max\left(\rho_0^2, \left(\frac{u_n}{\bar{X}_n}\right)^2\right) = \left(\frac{u_n}{\bar{X}_n}\right)^2$. Then our stopping rule is of the from

$$N \geq \left(\frac{u_N}{C\bar{X}_N}\right)^2 .$$

With no loss of generality, we assume $\mu > 0$. Now for $n = 1, 2, \ldots,$ define

$$Q_n = 1 + \sum_{k=1}^{n} (x_k - \mu)^2 .$$

For this case, the random variable

$$N(n^*) = \min\left(n\colon T_n \geq \frac{\mu}{\sigma}\sqrt{n^* Q_n}\right)$$

is well-defined and no smaller than N.

Now for $r = 1, 2, \ldots,$ define

$$R = \min\left(N(n^*),\ r\right) \quad \text{and} \quad B = \left(1 < N(n^*) \leq r\right).$$

We shall now apply Wald's theorem of cumulative sums to each of $T_R$, $Q_R - 1$, and $x_1^2 + x_2^2 + \ldots + x_R^2$ to obtain

$$\begin{aligned}
\mu E(R) &= E(T_R) \\
&= \int_{(R=1)} x_1 + \int_B T_{R-1} + \int_{(N(t)>r)} T_r + \int_B x_R \\
&\leq E|x_1| + \frac{\mu}{\sigma}\int_B \sqrt{n^*Q_{R-1}} + \frac{\mu}{\sigma}\int_{(N(t)>r)} \sqrt{n^*Q_r} + \int_B |x_R| \\
&\leq E^{\frac{1}{2}}(x_1^2) + \frac{\mu}{\sigma}E(\sqrt{n^*Q_R}) + E^{\frac{1}{2}}\left(\sum_{k=1}^{R} x_k^2\right) \\
&\leq \sqrt{\sigma^2 + \mu^2} + \frac{\mu}{\sigma}\sqrt{n^* + n^*\sigma^2 E(R)} + \sqrt{\sigma^2 + \mu^2}\, E^{\frac{1}{2}}(R) \\
&\leq \sqrt{\sigma^2 + \mu^2} + \frac{\mu}{\sigma}\sqrt{n^*} + \left(\mu\sqrt{n^*} + \sqrt{\sigma^2 + \mu^2}\right) E^{\frac{1}{2}}(R) .
\end{aligned}$$

This implies

$$E(R) \leq \sqrt{\rho^2 + 1} + \frac{1}{\sigma}\sqrt{n^*} + \left(\sqrt{n^*} + \sqrt{\rho^2 + 1}\right) E^{\frac{1}{2}}(R) .$$

Thus

$$E^{\frac{1}{2}}(R) \in \left\{x:\ x^2 - \left(\sqrt{n^*} + \sqrt{\rho^2 + 1}\right)x - \left(\frac{1}{\sigma}\sqrt{n^*} + \sqrt{\rho^2 + 1}\right) \leq 0\right\} .$$

From the set of x values above, we have

$$x \leq \sqrt{n^*} + o(\sqrt{n^*}) .$$

Hence

$$E(R) \leq \left(\sqrt{n^*} + o(\sqrt{n^*})\right)^2 .$$

Since $R \to N(t)$ as $r \to \infty$, $E(R) \to E(N(t))$ by the monotone convergence theorem. Thus

$$E[N(n^*)] \leq \left(\sqrt{n^*} + o(\sqrt{n^*})\right)^2$$

which implies

$$E\left(\frac{N(n^*)}{n^*}\right) \leq \left(1 + \frac{o\sqrt{n^*}}{\sqrt{n^*}}\right)^2 \quad .$$

Therefore,

$$\limsup_{n^* \to \infty} E\left(\frac{N(n^*)}{n^*}\right) \leq 1 \quad .$$

Since $N \leq N(n^*)$,

$$\limsup_{n^* \to \infty} E\left(\frac{N}{n^*}\right) \leq 1 \ .$$

Hence

$$\limsup_{C \to 0} E\left(\frac{N}{n^*}\right) \leq 1 \ .$$

From Cases (1) and (2), we have

$$\limsup_{C \to 0} E\left(\frac{N}{n^*}\right) \leq 1 \ . \tag{4.4}$$

(4.2) and (4.4) then yield the desired result,

$$\lim_{C \to 0} E\left(\frac{N}{n^*}\right) = 1 \ .$$

The following theorem states that the sequential process is risk efficient. The proof is the same as that for Theorem 1 with the exception that the upper bound on $\frac{n^*}{N}$ is now $\left(\frac{\sigma}{\mu\rho_0}\right)^2$. Thus the proof will not be given.

<u>Theorem 7</u>: $\lim\limits_{C \to 0} \dfrac{E(L_N)}{R_{n^*}(C)} = 1.$

Monte Carlo techniques were used to evaluate the sequential procedure, and the results for C = .3 and C = .1 are presented in Tables X and XI, respectively. When C = .3, the average random sample size tended to be noticeably smaller than n* and the $\widehat{CV}(\overline{X}_N)$ was greater than stated. Although there was some improvement when C = .1, $\widehat{CV}(\overline{X}_N)$ is still greater than the specified C.

TABLE X

NONPARAMETRIC, SEQUENTIAL ESTIMATION OF $\mu$ DESIGNED TO OBTAIN $CV(\bar{X}_N) = .3$ WHEN THE UNDERLYING DISTRIBUTION IS NEGATIVE BINOMIAL*

| $\mu$ | k | n* | $\bar{N}$ | $s_{\bar{N}}$ | $\bar{\bar{X}}_N$ | $s_{\bar{X}_N}$ | $\widehat{CV}(\bar{X}_N)$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 22.22 | 19.98 | .42 | 1.13 | .50 | .44 |
| | 2 | 16.67 | 15.60 | .36 | 1.13 | .46 | .40 |
| | 3 | 14.82 | 14.86 | .32 | 1.13 | .48 | .42 |
| | 4 | 13.89 | 13.29 | .31 | 1.17 | .47 | .41 |
| | 5 | 13.33 | 13.33 | .29 | 1.13 | .47 | .42 |
| 2 | 1 | 16.67 | 13.09 | .32 | 2.23 | .90 | .40 |
| | 2 | 11.11 | 9.44 | .25 | 2.25 | .88 | .39 |
| | 3 | 9.26 | 8.11 | .22 | 2.20 | .84 | .38 |
| | 4 | 8.33 | 7.70 | .20 | 2.18 | .82 | .38 |
| | 5 | 7.78 | 6.95 | .19 | 2.22 | .77 | .35 |
| 3 | 1 | 14.82 | 11.34 | .33 | 3.24 | 1.42 | .44 |
| | 2 | 9.26 | 7.28 | .20 | 3.31 | 1.37 | .41 |
| | 3 | 7.41 | 6.38 | .18 | 3.18 | 1.07 | .34 |
| | 4 | 6.48 | 5.26 | .15 | 3.30 | 1.13 | .34 |
| | 5 | 5.93 | 5.03 | .14 | 3.28 | 1.15 | .35 |
| 4 | 1 | 13.89 | 10.34 | .30 | 4.27 | 1.99 | .47 |
| | 2 | 8.33 | 6.35 | .19 | 4.16 | 1.66 | .40 |
| | 3 | 6.48 | 4.99 | .14 | 4.15 | 1.60 | .38 |
| | 4 | 5.56 | 4.44 | .13 | 4.24 | 1.58 | .37 |
| | 5 | 5.00 | 4.15 | .11 | 4.20 | 1.39 | .33 |
| 5 | 1 | 13.33 | 9.66 | .28 | 5.25 | 2.41 | .46 |
| | 2 | 7.78 | 5.58 | .17 | 5.05 | 1.98 | .39 |
| | 3 | 5.93 | 4.68 | .14 | 5.28 | 1.96 | .37 |
| | 4 | 5.00 | 3.99 | .11 | 5.32 | 1.89 | .36 |
| | 5 | 4.44 | 3.71 | .10 | 5.32 | 1.95 | .37 |

* Each entry is based on 500 simulations.

TABLE XI

NONPARAMETRIC, SEQUENTIAL ESTIMATION OF μ DESIGNED TO OBTAIN $CV(\bar{X}_N) = .1$ WHEN THE UNDERLYING DISTRIBUTION IS NEGATIVE BINOMIAL*

| $\mu$ | k | n* | $\bar{N}$ | $s_{\bar{N}}$ | $\bar{\bar{X}}_N$ | $s_{\bar{X}_N}$ | $\widehat{CV}(\bar{X}_N)$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 200.00 | 197.42 | 1.29 | 1.00 | .10 | .10 |
| | 2 | 150.00 | 149.31 | 1.03 | 1.01 | .10 | .10 |
| | 3 | 133.33 | 133.33 | .98 | 1.01 | .11 | .11 |
| | 4 | 125.00 | 125.67 | .92 | 1.01 | .10 | .10 |
| | 5 | 120.00 | 118.68 | .90 | 1.02 | .11 | .10 |
| 2 | 1 | 150.00 | 145.79 | 1.24 | 2.03 | .40 | .20 |
| | 2 | 100.00 | 97.49 | .87 | 2.02 | .28 | .14 |
| | 3 | 83.33 | 81.18 | .75 | 2.02 | .33 | .16 |
| | 4 | 75.00 | 72.61 | .76 | 2.04 | .27 | .13 |
| | 5 | 70.00 | 69.43 | .70 | 2.03 | .30 | .15 |
| 3 | 1 | 133.33 | 126.49 | 1.20 | 3.04 | .54 | .18 |
| | 2 | 83.33 | 79.60 | .87 | 3.08 | .62 | .20 |
| | 3 | 66.67 | 63.09 | .73 | 3.09 | .44 | .14 |
| | 4 | 58.33 | 55.75 | .63 | 3.05 | .46 | .15 |
| | 5 | 53.33 | 50.68 | .65 | 3.07 | .48 | .16 |
| 4 | 1 | 125.00 | 118.19 | 1.29 | 4.15 | .96 | .23 |
| | 2 | 75.00 | 70.29 | .86 | 4.10 | .75 | .18 |
| | 3 | 58.33 | 53.76 | .85 | 4.14 | .90 | .22 |
| | 4 | 50.00 | 45.78 | .70 | 4.19 | .89 | .21 |
| | 5 | 45.00 | 40.94 | .65 | 4.16 | .80 | .19 |
| 5 | 1 | 120.00 | 112.37 | 1.19 | 5.10 | .98 | .19 |
| | 2 | 70.00 | 63.66 | .95 | 5.17 | .90 | .17 |
| | 3 | 53.33 | 47.91 | .79 | 5.16 | .92 | .18 |
| | 4 | 45.00 | 40.18 | .67 | 5.20 | .97 | .19 |
| | 5 | 40.00 | 33.60 | .67 | 5.18 | .90 | .17 |

* Each entry is based on 500 simulations.

Although we do not have an analytical proof of its existence, inspection of the tables shows that $\overline{X}_N$ has a definite positive bias under this nonparametric procedure. It is greater than what was observed under the corresponding method based on the negative binomial distribution.

Viewing these tables in light of the results presented in Table III, we believe that if a fairly reliable estimate of k is available, it would be better to use that than to turn too quickly to the nonparametric approach presented here.

# CHAPTER V

## ESTIMATION OF THE PARAMETER k

We shall now focus our attention on the estimation of k for the two-parameter negative binomial distribution. The distribution will be examined for completeness. We shall present a multistage procedure for estimating k and compare it to the method of moments (MME) and maximum likelihood (MLE) estimators.

### Complete Sufficient Statistic for Samples of Fixed Size n

A common procedure in statistical inference is to determine the minimal sufficient statistic for a family of distributions and examine the family for completeness. If it is complete, then we are ready to search for minimum variance unbiased estimators of the parameters. However, we shall prove in the next theorem that there is no complete sufficient statistic for the negative binomial family of distributions.

Theorem 8: The order statistics, $X_{(1)}$, $X_{(2)}$, ..., $X_{(n)}$, are minimal sufficient but not complete for the negative binomial family of distributions when the sample size exceeds n = 3.

Proof: We shall use the 1950 Lehmann-Scheffé theorem (21) to determine the minimal sufficient statistic. For any point $\underset{\sim}{x}^{o}$, $D(\underset{\sim}{x}^{o})$ is defined as the set of all points $\underset{\sim}{x}$ for which there exists a function $k(\underset{\sim}{x}, \underset{\sim}{x}^{o}) \neq 0$, not depending on μ or k, and such that $P_{\mu,k}(\underset{\sim}{x}) = k(\underset{\sim}{x}, \underset{\sim}{x}^{o})P_{\mu,k}(\underset{\sim}{x}^{0})$

for all $\mu$, k in the parameter space; that is,

$$D(\underset{\sim}{x}^o) = \{\underset{\sim}{x} \mid P_{\mu,k}(\underset{\sim}{x}) = k(\underset{\sim}{x}, \underset{\sim}{x}^o)P_{\mu,k}(\underset{\sim}{x}^o) \text{ for all } \mu, k\} .$$

Roughly speaking, $D(\underset{\sim}{x}^o)$ consists of all x for which the ratio $\dfrac{P_{\mu,k}(\underset{\sim}{x})}{P_{\mu,k}(\underset{\sim}{x}^o)}$ is independent of $\mu$ and k. The minimal sufficient statistic, T, is the statistic of lowest dimension such that $T(\underset{\sim}{x}) = T(\underset{\sim}{x}^o)$ if $\underset{\sim}{x} \in D(x^o)$ and $T(x) \neq T(x^o)$ if $\underset{\sim}{x} \notin D(x^o)$ . We have

$$\frac{P(\underset{\sim}{x})}{P(\underset{\sim}{x}^o)} = \frac{\left(\frac{k}{k+\mu}\right)^{nk}\left(\frac{\mu}{\mu+k}\right)^{\sum_{i=1}^{n} x_i}\left(\frac{1}{(k-1)!}\right)^n \prod_{i=1}^{n} \frac{(x_i+k-1)!}{x_i!}}{\left(\frac{k}{k+\mu}\right)^{nk}\left(\frac{\mu}{\mu+k}\right)^{\sum_{i=1}^{n} x_i^o}\left(\frac{1}{(k-1)!}\right)^n \prod_{i=1}^{n} \frac{(x_i^o+k-1)!}{x_i^o!}}$$

$$= \left(\frac{\mu}{\mu+k}\right)^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^o} \prod_{i=1}^{n} \frac{x_i^o!(x_i+k-1)!}{x_i!(x_i^o+k-1)!} \qquad (5.1)$$

$$\overset{?}{=} k(\underset{\sim}{x}, \underset{\sim}{x}^o) .$$

In order to obtain $k(\underset{\sim}{x}, \underset{\sim}{x}^o)$ independent of $\mu$ and k in (5.1), we must have

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i^o$$

and

$$\prod_{i=1}^{n} \frac{x_i^o!(x_i+k-1)!}{x_i!(x_i^o+k-1)!} = c \qquad (5.2)$$

for some constant c. We find that for equality to hold in (5.2), there must be an $x_j$, j = 1, ..., n, corresponding to each $x_i^o$ , i = 1, ..., n, such that $x_j = x_i^o$. Thus the order statistics of $\underset{\sim}{x}$ must equal those of

$\underset{\sim}{x}^o$ for $\underset{\sim}{x} \in D(\underset{\sim}{x}^o)$. Therefore, the order statistics form the minimal sufficient statistic.

We shall now prove the order statistics are not complete. To do this, we shall show that there exists a nontrivial function $g(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ such that $E\left(g(x_{(1)}, x_{(2)}, \ldots, x_{(n)})\right) = 0$. That is,

$$0 = \sum_{j=0}^{\infty} g(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) \frac{n!}{\pi_j m_j} \frac{\pi_j \{(k+j-1)!\}^{m_j}}{\{(k-1)!\}^n \pi_j \{j!\}^{m_j}} \left(\frac{k}{\mu+k}\right)^{nk} .$$

$$\left(\frac{\mu}{\mu+k}\right)^{\sum_{i=1}^{n} x_i}$$

where $m_j$ represents the number of times a j was observed in a sample of size n. To simplify notation, let $p = \frac{k}{\mu+k}$ and $q = \frac{\mu}{\mu+k}$. Then

$$\begin{aligned} 0 = \frac{n!p^{nk}}{\{(k-1)!\}^n} \Bigg[ & \frac{\{(k-1)!\}^n}{n!} g(0, \ldots, 0)\, q^o \\ & + \frac{\{(k-1)!\}^{n-1} k!}{(n-1)!} g(0, \ldots, 0, 1) q^1 \\ & + \left( \frac{\{(k-1)!\}^{n-1} (k+1)!}{(n-1)!2!} g(0, \ldots, 0, 2) \right. \\ & \left. + \frac{\{(k-1)!\}^{n-2} (k!)^2}{(n-2)!2!} g(0, \ldots, 0, 1, 1) \right) q^2 \qquad (5.3) \\ & + \left( \frac{\{(k-1)!\}^{n-1} (k+2)!}{(n-1)!3!} g(0, \ldots, 0, 3) \right. \\ & + \frac{\{(k-1)!\}^{n-2} k!(k+1)!}{(n-2)!2!} g(0, \ldots, 0, 1, 2) \\ & \left. + \frac{\{(k-1)!\}^{n-3} (k!)^3}{(n-3)!3!} g(0, \ldots, 0, 1, 1, 1) \right) q^3 + \ldots \Bigg] . \end{aligned}$$

Viewing (5.3) as a polynomial in q, equality can hold if and only the coefficients of each of the powers of q are zero. After considerable algebra, we find the coefficients of $q^0$, $q^1$, $q^2$, and $q^3$ are zero.

Consider now the coefficient of $q^c$ where $c \le n$ and $n \ge 4$. We have

$$\frac{\{(k-1)!\}^{n-1}(k+c-1)!}{(n-1)!c!} g(0, \ldots, 0, c)$$

$$+ \frac{\{(k-1)!\}^{n-2}k!(k+c-2)!}{(n-2)!(c-1)!} g(0, \ldots, 0, 1, c-1)$$

$$+ \frac{\{(k-1)!\}^{n-2}(k+1)!(k+c-3)!}{(n-2)!2!(c-2)!} g(0, \ldots, 0, 2, c-2)$$

$$+ \frac{\{(k-1)!\}^{n-2}(k+2)!(k+c-4)!}{(n-2)!3!(c-3)!} g(0, \ldots, 0, 3, c-3) + \ldots$$

$$+ \frac{\{(k-1)!\}^{n-2}(k+[\frac{c}{2}]-2)!(k+c-[\frac{c}{2}])!}{(n-2)!([\frac{c}{2}]-1)!(c-[\frac{c}{2}]+1)!} g(0, \ldots, 0, [\tfrac{c}{2}]-1, c-[\tfrac{c}{2}]+1)$$

$$+ \ldots + \frac{\{(k-1)!\}^{n-m}(k!)^{m-1}(k+1)!(k+c-m)!}{(n-m)!(m-1)!(c-m+1)!} g(0, \ldots, 0, 1, \ldots, 1, c-m+1)$$

$$+ \frac{\{(k-1)!\}^{n-m}(k!)^{m-2}(k+1)!(k+c-m-1)!}{(n-m)!(m-2)!2!(c-m)!} g(0, \ldots, 0, 1, \ldots, 1, 2, c-m)$$

$$+ \ldots + \frac{\{(k-1)!\}^{n-m}\{(k+[\frac{c}{m}]-1)!\}^{m-1}(k+c-(m-1)[\frac{c}{m}]-1)!}{(n-m)!(m-1)!\{([\frac{c}{m}])!\}^{m-1}(c-(m-1)[\frac{c}{m}])!} \cdot$$

$$g(0, \ldots, 0, [\tfrac{c}{m}], \ldots, [\tfrac{c}{m}], c-(m-1)[\tfrac{c}{m}])$$

$$+ \ldots + \frac{\{(k-1)!\}^{n-c+1}(k!)^{c-2}(k+1)!}{(n-c+1)!(c-2)!2!} g(0, \ldots, 0, 1, \ldots, 1, 2)$$

$$+ \frac{\{(k-1)!\}^{n-c}(k!)^{c}}{(n-c)!c!} g(0, \ldots, 0, 1, \ldots, 1) = 0 .$$

We can now factor out $\{(k-1)!\}^{n-1}k!$ Considering the resulting equation as a polynomial in k, we have $g(0, \ldots, 0, c) = 0$. We can then remove a common k. Hence we obtain

$$\frac{(k+c-2)\ \ldots\ (k+1)}{(n-2)!(c-1)!}\ g(0,\ \ldots,\ 0,\ 1,\ c-1)$$

$$+\ \frac{(k+1)(k+c-3)\ \ldots\ (k+1)}{(n-2)!2!(c-2)!}\ g(0,\ \ldots,\ 0,\ 2,\ c-2)$$

$$+\ \frac{(k+2)(k+1)(k+c-4)\ \ldots\ (k+1)}{(n-2)!3!(c-3)!}\ g(0,\ \ldots,\ 0,\ 3,\ c-3)$$

$$+\ \ldots\ +\ \frac{(k+[\frac{c}{2}]-2)\ \ldots\ (k+1)(k+c-[\frac{c}{2}])\ \ldots\ (k+1)}{(n-2)!([\frac{c}{2}]-1)!(c-[\frac{c}{2}]+1)!}\ .$$

$$g(0,\ \ldots,\ 0,\ [\tfrac{c}{2}]-1,\ c-[\tfrac{c}{2}]+1)$$

$$+\ \ldots\ +\ \frac{k^{m-2}(k+c-m)\ \ldots\ (k+1)}{(n-m)!(m-1)!(c-m+1)!}\ g(0,\ \ldots,\ 0,\ 1,\ \ldots,\ 1,\ c-m+1)$$

$$+\ \frac{k^{m-3}(k+1)k(k+c-m-1)\ \ldots\ (k+1)}{(n-m)!(m-2)!2!(c-m)!}\ g(0,\ \ldots,\ 0,\ 1,\ \ldots,\ 1,2,\ c-m)$$

$$+\ \ldots\ +\ \frac{k^{m-2}\{(k+[\frac{c}{m}]-1)\ \ldots\ (k+1)\}^{m-1}(k+c-(m-1)[\frac{c}{m}]-1)\ \ldots\ (k+1)}{(n-m)!(m-1)!\{([\frac{c}{m}])!\}^{m-1}(c-(m-1)[\frac{c}{m}])!}\ .$$

$$g(0,\ \ldots,\ 0,\ [\tfrac{c}{m}],\ \ldots,\ [\tfrac{c}{m}],\ c-(m-1)[\tfrac{c}{m}])$$

$$+\ \ldots\ +\ \frac{k^{c-3}(k+1)}{(n-c+1)!(c-2)!2!}\ g(0,\ \ldots,\ 0,\ 1,\ \ldots,\ 1,\ 2)$$

$$+\ \frac{k^{c-2}}{(n-c)!c!}\ g(0,\ \ldots,\ 0,\ 1,\ \ldots,\ 1)\ =\ 0\ .$$

Here we note that from the polynomial in k, we obtain c - 1 constraints on the g's. If there are more than (c -1) g's, then there will be an infinite number of solutions. As an example, we shall consider the case of c = 4. From the general expansion, we have

$$\frac{(k+2)(k+1)}{(n-2)!3!}\ g(0,\ \ldots,\ 0,\ 1,\ 3)\ +\ \frac{(k+1)(k+1)}{(k-2)!(2!)^3}\ g(0,\ldots,\ 0,\ 2,\ 2)$$

$$+ \frac{k(k+1)}{(n-3)!2!2!} g(0, \ldots, 0, 1, 1, 2)$$

$$+ \frac{k^2}{(n-4)!4!} g(0, \ldots, 0, 1, \ldots, 1) = 0.$$

The coefficients of the polynomials in k produce three constraints on the four g's. Hence we have a homogeneous system of three equations in four unknowns, and there are an infinite number of ways that the function g can be defined that will satisfy these equations. Therefore, there exists a nontrivial function $g(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ such that the expectation of g is zero. This proves the order statistics are not complete.

## Method of Moments and Maximum Likelihood Estimation of k

For a sample of fixed size n from a negative binomial distribution with parameters $\mu$ and k, the likelihood function is

$$L(\mu, k) = \prod_{i=1}^{n} \left[ \frac{(k+x_i-1)!}{x_i!k!} \left(\frac{k}{k+\mu}\right)^k \left(\frac{\mu}{k+\mu}\right)^x \right]$$

$$= \prod_{i=1}^{n} \left[ \frac{(k+x_i-1)!}{x_i!} \right] \frac{1}{(k!)^n} \left(\frac{k}{\mu+k}\right)^{nk} \left(\frac{\mu}{\mu+k}\right)^{\sum_{i=1}^{n} x_i} .$$

Hence the natural logarithm of the likelihood function is

$$\ell n\ L = \sum_{i=1}^{n} \ell n(k+x_i-1)! - \sum_{i=1}^{n} \ell n\ x_i! - n\ \ell n\ k! + nk\ \ell n\ k$$

$$+ (\ell n\ \mu) \sum_{i=1}^{n} x_i - \left(nk + \sum_{i=1}^{n} x_i\right) \ell n(k+\mu) .$$

As noted in Chapter II, no closed form solution exists for the maximum likelihood estimate of k which is the root of the following equation in $\hat{k}$:

$$n\,\ln\left(1+\frac{\overline{X}_n}{\hat{k}}\right) = \sum_{j=1}^{\infty} m_j\left(\frac{1}{\hat{k}} + \frac{1}{\hat{k}+1} + \ldots + \frac{1}{\hat{k}+j-1}\right) . \qquad (5.4)$$

Although it has been proven that there is at least one root of (5.4) when $s^2 > \overline{X}$, we do not know if it is unique. Furthermore, there has been no proof that a solution does not exist if $s^2 < \overline{X}$.

In order to better understand the likelihood function, we plotted some contours of the natural logarithm of the likelihood function using S.A.S. Four of these are shown in Figures 5-8. Although we have viewed only a few plots, we note there is a basic similarity in the contours. The dominant characteristic is the appearance of long, narrow ridges. The narrowness of the ridges indicates $\mu$ can be estimated with precision, using $\overline{X}$, the MLE of $\mu$. However, we believe the long length of the ridges is an indication that maximum likelihhod estimation of k is not precise.

We computed the estimated biases of the MME and MLE of k in much the same manner as Pieters, Gates, Matis, and Sterling (31). In addition, we estimated the standard deviation of the estimates and combined the estimates of the bias and standard deviation in estimating the mean square error (MSE). Results for fixed sample sizes of 50, 100, and 200 are presented in Tables XII, XIII, and XIV.

Upon inspection of the tables, it appears that there is less bias and a smaller standard deviation under MLE than under MME. Even though the estimated mean square error is smaller for MLE, it is still large. It is interesting to note that for a fixed $\mu$, the estimates of the bias, standard deviation, and mean square error all tend to increase as k increases. For a fixed k, they all tend to decrease as $\mu$ increases. Thus estimation is most difficult when $\mu$ is small and k is large.

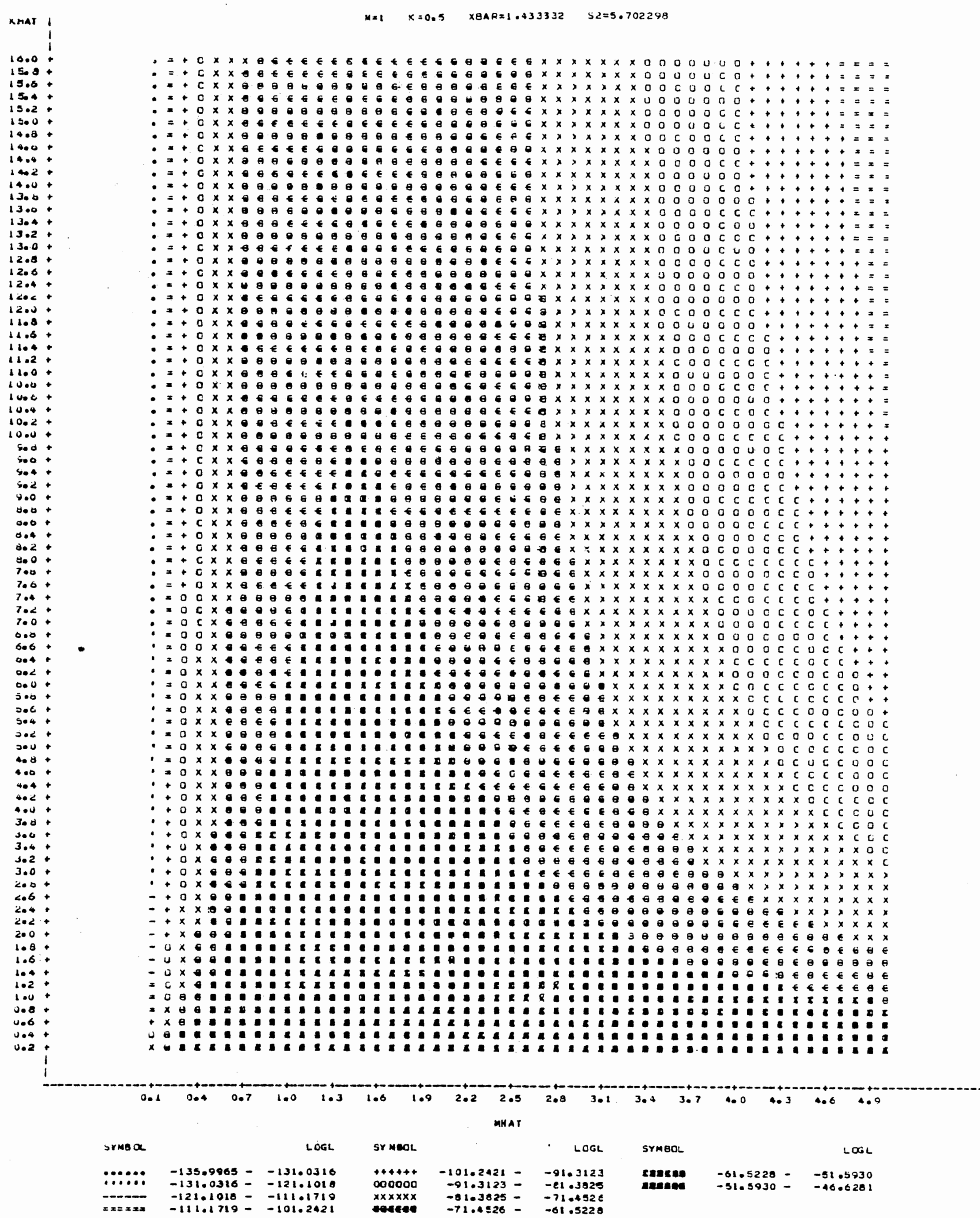



Figure 5. Contour of the Maximum Likelihood Function Based on a Sample of Size 30 Drawn from NB(1, .5) and Having a Maximum Likelihood Estimate $\hat{k}$ = .3186

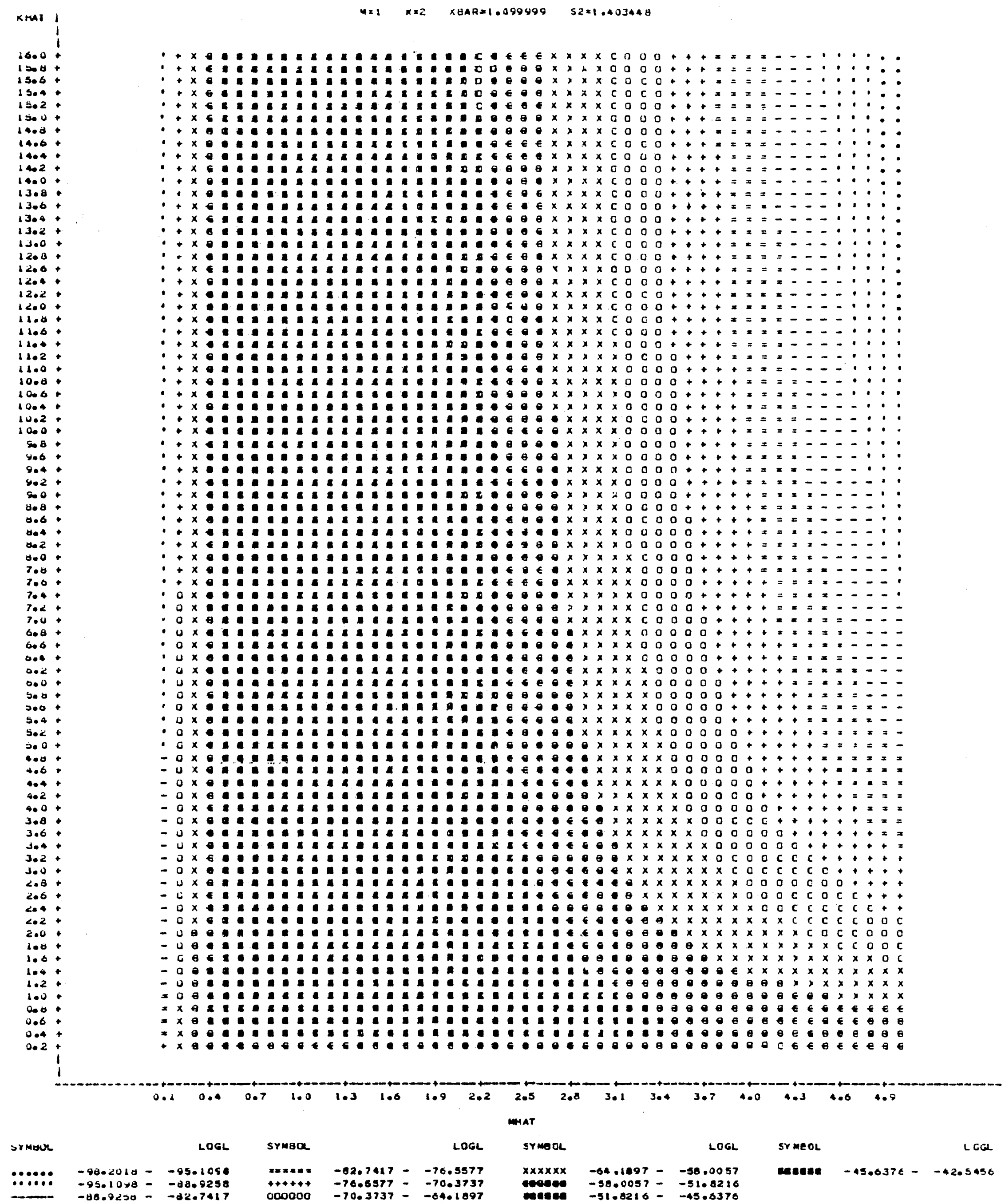
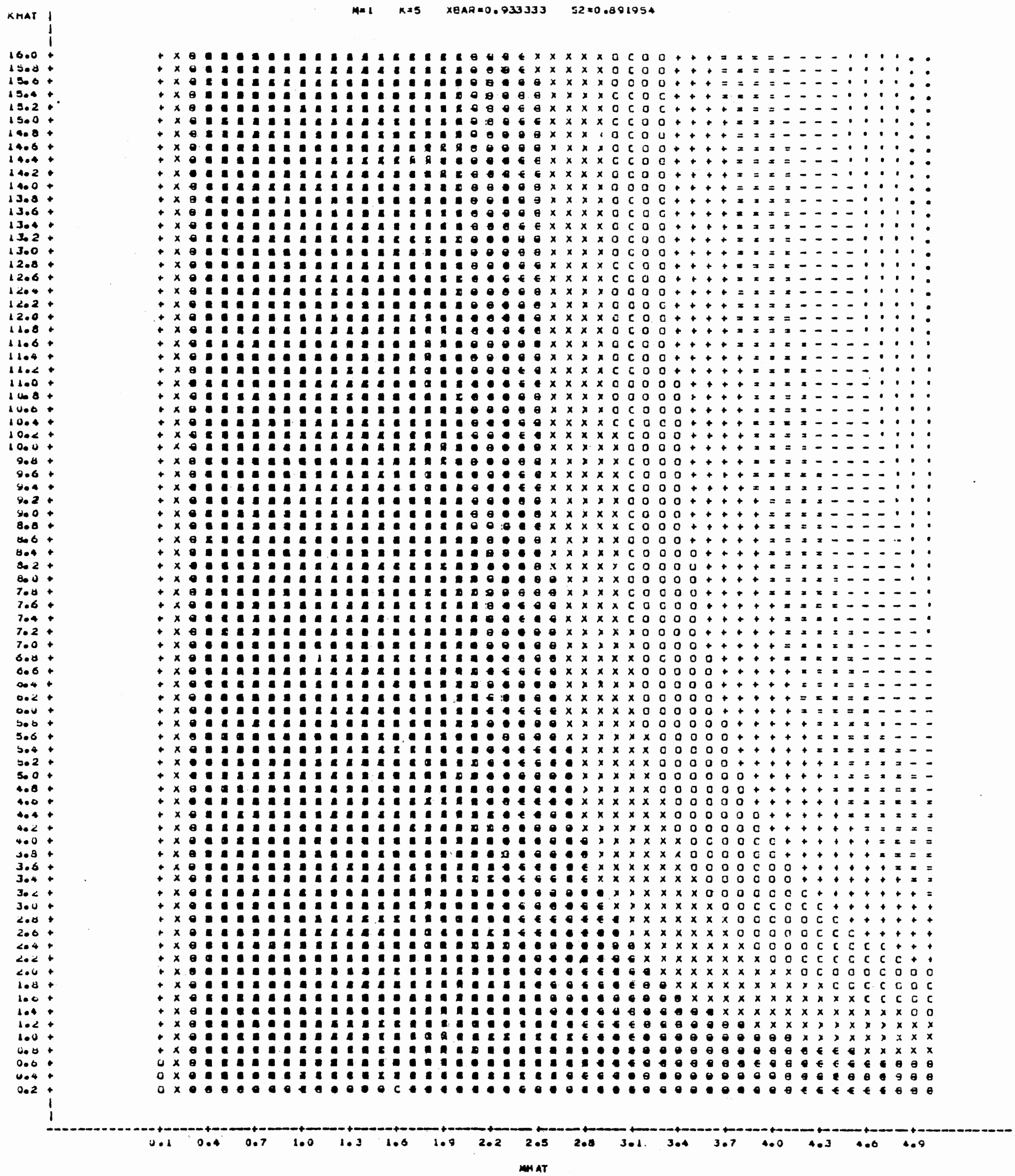
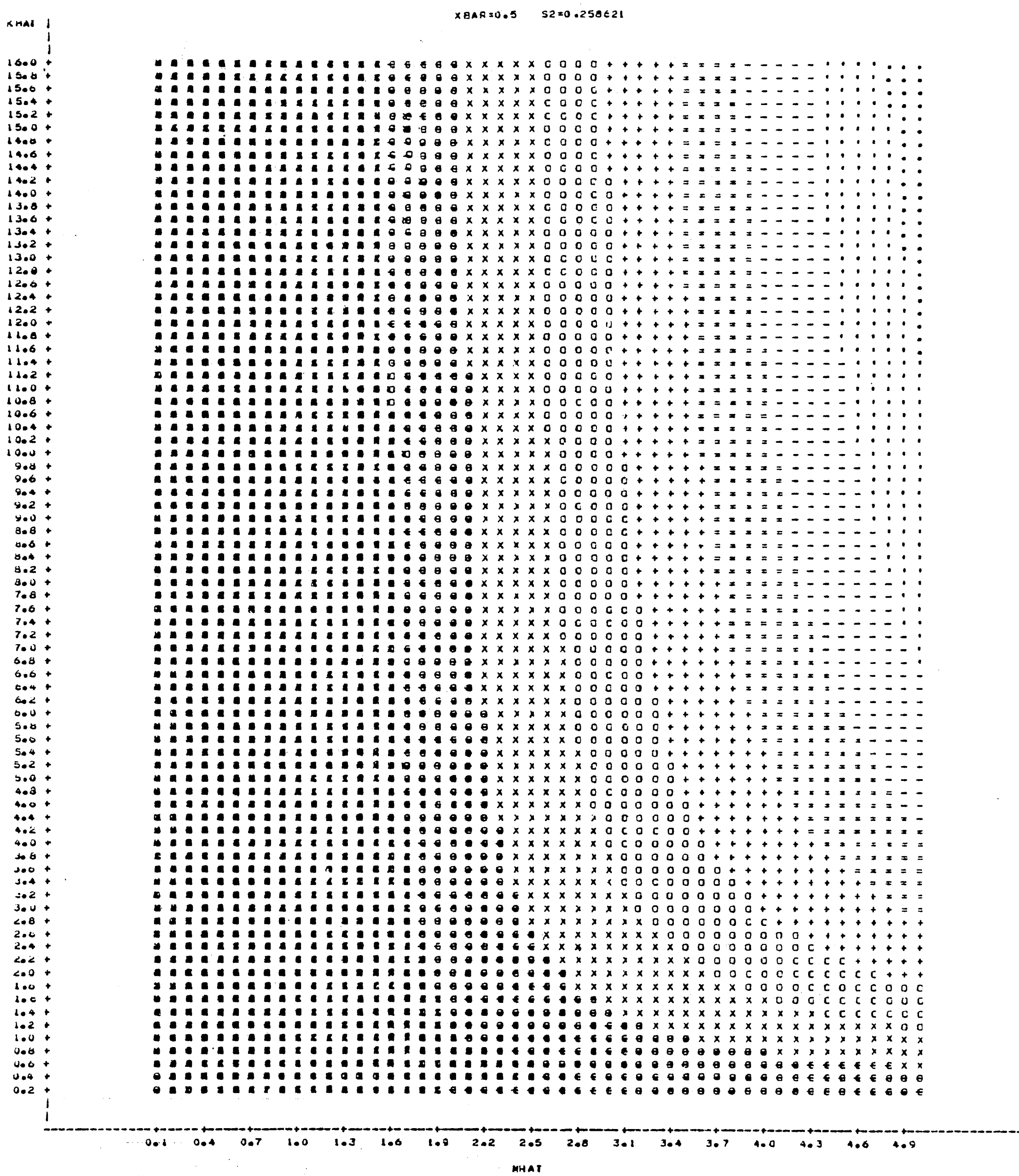



Figure 6. Contour of the Maximum Likelihood Function Based on a Sample of Size 30 Drawn from NB(1, 2) and Having a Maximum Likelihood Estimate $\hat{k}$ = 3.6469




| SYMBOL | LOGL | SYMBOL | LOGL | SYMBOL | LOGL |
|---|---|---|---|---|---|
| •••••• | -100.0999 - -96.6372 | ++++++ | -75.8613 - -68.9360 | ■■■■■■ | -48.1600 - -41.2347 |
| '''''' | -96.6372 - -89.7119 | OOOOOO | -68.9360 - -62.0106 | ■■■■■■ | -41.2347 - -37.7720 |
| ------ | -89.7119 - -82.7866 | XXXXXX | -62.0106 - -55.0853 | | |
| ====== | -82.7866 - -75.8613 | θθθθθθ | -55.0853 - -48.1600 | | |

Figure 7. Contour of the Maximum Likelihood Function Based on a Sample of Size 30 Drawn from NB(1, 5) and Having No Maximum Likelihood Estimate of k




| SYMBOL | LOGL | SYMBOL | LOGL | SYMBOL | LOGL |
|---|---|---|---|---|---|
| ...... | -110.4656 - -105.7525 | ++++++ | -77.4737 - -68.0474 | ■■■■■■ | -39.7686 - -30.3423 |
| '''''' | -105.7525 - -96.3262 | OOOOOO | -68.0474 - -58.6211 | ■■■■■■ | -30.3423 - -25.6292 |
| ------ | -96.3262 - -86.8999 | XXXXXX | -58.6211 - -49.1949 | | |
| ====== | -86.8999 - -77.4737 | @@@@@@ | -49.1949 - -39.7686 | | |

Figure 8. Contour of the Maximum Likelihood Function Based on a Sample of Size 30 Consisting of Fifteen Zeroes and Fifteen Ones and Having No Maximum Likelihood Estimate of k

TABLE XII

METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD ESTIMATION OF k BASED ON SAMPLES OF SIZE 50*

| $\mu$ | k | Method of Moments | | | | Maximum Likelihood | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\bar{\hat{k}}$ | $\widehat{Bias}$ | $s_{\hat{k}}$ | $\widehat{MSE}$ | $\bar{\hat{k}}$ | $\widehat{Bias}$ | $s_{\hat{k}}$ | $\widehat{MSE}$ |
| 1 | 1 | 1.967 | .967 | 8.597 | 74.850 | 1.549 | .549 | 2.616 | 7.146 |
| | 2 | 4.946 | 2.946 | 17.790 | 325.154 | 3.622 | 1.622 | 5.406 | 31.849 |
| | 3 | 8.329 | 5.329 | 25.012 | 653.999 | 5.307 | 2.307 | 7.249 | 57.874 |
| | 4 | 10.195 | 6.195 | 29.089 | 884.532 | 6.363 | 2.363 | 8.529 | 78.318 |
| | 5 | 12.356 | 7.356 | 32.788 | 1129.165 | 7.078 | 2.078 | 9.016 | 85.600 |
| 2 | 1 | 1.240 | .240 | .630 | .454 | 1.194 | .194 | .680 | .500 |
| | 2 | 2.833 | .833 | 3.359 | 11.980 | 2.845 | .845 | 3.580 | 13.533 |
| | 3 | 5.164 | 2.164 | 11.283 | 131.978 | 5.085 | 2.085 | 7.363 | 58.556 |
| | 4 | 8.293 | 4.293 | 22.147 | 508.924 | 7.197 | 3.197 | 9.955 | 109.319 |
| | 5 | 12.771 | 7.771 | 41.584 | 1789.587 | 8.502 | 3.502 | 10.468 | 121.851 |
| 3 | 1 | 1.173 | .173 | .451 | .234 | 1.134 | .134 | .400 | .178 |
| | 2 | 2.458 | .458 | 1.934 | 3.949 | 2.331 | .331 | 1.145 | 1.421 |
| | 3 | 4.689 | 1.689 | 14.853 | 223.480 | 4.069 | 1.069 | 3.529 | 13.597 |
| | 4 | 6.517 | 2.517 | 12.408 | 160.295 | 5.996 | 1.996 | 6.648 | 48.179 |
| | 5 | 10.242 | 5.242 | 31.370 | 1011.574 | 8.011 | 3.011 | 9.345 | 96.390 |
| 4 | 1 | 1.137 | .137 | .406 | .184 | 1.101 | .101 | .322 | .114 |
| | 2 | 2.369 | .369 | 1.081 | 1.305 | 2.316 | .316 | .995 | 1.091 |
| | 3 | 3.707 | .707 | 2.452 | 6.511 | 3.612 | .612 | 1.777 | 3.532 |
| | 4 | 7.493 | 3.493 | 52.857 | 2806.050 | 5.502 | 1.502 | 5.220 | 29.504 |
| | 5 | 7.652 | 2.652 | 16.090 | 265.930 | 6.956 | 1.956 | 6.457 | 45.525 |
| 5 | 1 | 1.141 | .141 | .353 | .145 | 1.088 | .088 | .309 | .103 |
| | 2 | 2.319 | .319 | .892 | .897 | 2.254 | .254 | .815 | .728 |
| | 3 | 3.477 | .477 | 1.886 | 3.783 | 3.491 | .491 | 1.567 | 2.698 |
| | 4 | 5.083 | 1.083 | 4.854 | 24.735 | 5.020 | 1.020 | 3.489 | 13.217 |
| | 5 | 7.236 | 2.236 | 18.397 | 343.456 | 6.511 | 1.511 | 4.791 | 25.238 |

* Each entry is based on 1000 simulations.

TABLE XIII

METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD ESTIMATION OF k BASED ON SAMPLES OF SIZE 100*

| $\mu$ | k | Method of Moments | | | | Maximum Likelihood | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\bar{\hat{k}}$ | $\widehat{\text{Bias}}$ | $s_{\hat{k}}$ | $\widehat{\text{MSE}}$ | $\bar{\hat{k}}$ | $\widehat{\text{Bias}}$ | $s_{\hat{k}}$ | $\widehat{\text{MSE}}$ |
| 1 | 1 | 1.189 | .189 | .659 | .470 | 1.168 | .168 | .640 | .438 |
| | 2 | 3.111 | 1.111 | 7.325 | 54.889 | 3.212 | 1.212 | 5.230 | 28.817 |
| | 3 | 9.354 | 6.354 | 52.466 | 2793.056 | 5.098 | 2.098 | 7.169 | 55.792 |
| | 4 | 9.432 | 5.432 | 25.021 | 655.542 | 6.930 | 2.930 | 9.472 | 98.303 |
| | 5 | 12.487 | 7.487 | 44.594 | 2044.662 | 7.986 | 2.986 | 11.023 | 130.430 |
| 2 | 1 | 1.107 | .107 | .321 | .115 | 1.076 | .076 | .277 | .082 |
| | 2 | 2.282 | .282 | .937 | .957 | 2.265 | .265 | .853 | .797 |
| | 3 | 3.766 | .766 | 3.057 | 9.930 | 3.932 | .932 | 4.066 | 17.399 |
| | 4 | 5.350 | 1.350 | 4.550 | 22.530 | 5.956 | 1.956 | 7.917 | 66.499 |
| | 5 | 8.569 | 3.569 | 16.267 | 277.361 | 7.785 | 2.785 | 8.614 | 81.953 |
| 3 | 1 | 1.079 | .079 | .279 | .084 | 1.050 | .050 | .229 | .055 |
| | 2 | 2.221 | .221 | .707 | .549 | 2.198 | .198 | .707 | .540 |
| | 3 | 3.481 | .481 | 1.483 | 2.432 | 3.370 | .370 | 1.488 | 2.349 |
| | 4 | 4.796 | .796 | 2.557 | 7.173 | 4.821 | .821 | 3.478 | 12.774 |
| | 5 | 6.185 | 1.185 | 4.468 | 21.366 | 6.376 | 1.376 | 4.852 | 25.433 |
| 4 | 1 | 1.080 | .080 | .261 | .074 | 1.050 | .050 | .213 | .048 |
| | 2 | 2.164 | .164 | .619 | .409 | 2.123 | .123 | .508 | .274 |
| | 3 | 3.337 | .337 | 1.174 | 1.491 | 3.329 | .329 | 1.012 | 1.133 |
| | 4 | 4.583 | .583 | 1.822 | 3.660 | 4.610 | .610 | 2.410 | 6.180 |
| | 5 | 6.024 | 1.024 | 3.451 | 12.961 | 5.996 | .996 | 2.998 | 9.979 |
| 5 | 1 | 1.090 | .090 | .242 | .067 | 1.051 | .051 | .196 | .041 |
| | 2 | 2.125 | .125 | .529 | .295 | 2.146 | .146 | .494 | .265 |
| | 3 | 3.263 | .263 | .865 | .817 | 3.210 | .210 | .821 | .717 |
| | 4 | 4.339 | .339 | 1.382 | 2.025 | 4.347 | .347 | 1.420 | 2.136 |
| | 5 | 5.539 | .539 | 2.113 | 4.756 | 5.554 | .554 | 2.036 | 4.453 |

* Each entry is the result of 1000 simulations.

TABLE XIV

METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD ESTIMATION OF k BASED ON SAMPLES OF SIZE 200*

| $\mu$ | k | Method of Moments | | | | Maximum Likelihood | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\overline{\hat{k}}$ | $\widehat{\text{Bias}}$ | $s_{\hat{k}}$ | $\widehat{\text{MSE}}$ | $\overline{\hat{k}}$ | $\widehat{\text{Bias}}$ | $s_{\hat{k}}$ | $\widehat{\text{MSE}}$ |
| 1 | 1 | 1.097 | .097 | .374 | .149 | 1.060 | .060 | .297 | .092 |
| | 2 | 2.354 | .354 | 1.310 | 1.840 | 2.400 | .400 | 1.473 | 2.329 |
| | 3 | 4.364 | 1.364 | 9.200 | 86.498 | 4.190 | 1.190 | 5.238 | 28.851 |
| | 4 | 9.927 | 5.927 | 64.621 | 4211.008 | 6.428 | 2.428 | 7.846 | 67.463 |
| | 5 | 12.586 | 7.586 | 40.730 | 1716.490 | 8.119 | 3.119 | 9.554 | 101.010 |
| 2 | 1 | 1.057 | .057 | .218 | .051 | 1.044 | .044 | .195 | .040 |
| | 2 | 2.139 | .139 | .599 | .378 | 2.125 | .125 | .539 | .306 |
| | 3 | 3.309 | .309 | 1.308 | 1.806 | 3.308 | .308 | 1.153 | 1.425 |
| | 4 | 4.628 | .628 | 3.298 | 11.271 | 4.729 | .729 | 3.204 | 10.800 |
| | 5 | 6.489 | 1.489 | 15.041 | 228.437 | 5.940 | .940 | 3.307 | 11.821 |
| 3 | 1 | 1.035 | .035 | .184 | .035 | 1.027 | .027 | .155 | .025 |
| | 2 | 2.100 | .100 | .445 | .208 | 2.086 | .086 | .385 | .156 |
| | 3 | 3.166 | .166 | .788 | .648 | 3.170 | .170 | .757 | .603 |
| | 4 | 4.346 | .346 | 1.280 | 1.758 | 4.312 | .312 | 1.220 | 1.586 |
| | 5 | 5.507 | .507 | 1.914 | 3.919 | 5.554 | .554 | 2.561 | 6.868 |
| 4 | 1 | 1.039 | .039 | .179 | .033 | 1.020 | .020 | .137 | .019 |
| | 2 | 2.077 | .077 | .395 | .162 | 2.051 | .051 | .333 | .113 |
| | 3 | 3.149 | .149 | .706 | .521 | 3.128 | .128 | .613 | .392 |
| | 4 | 4.219 | .219 | 1.022 | 1.093 | 4.223 | .223 | .957 | .965 |
| | 5 | 5.368 | .368 | 1.594 | 2.675 | 5.509 | .509 | 1.975 | 4.162 |
| 5 | 1 | 1.039 | .039 | .158 | .026 | 1.028 | .028 | .125 | .017 |
| | 2 | 2.081 | .081 | .377 | .148 | 2.056 | .056 | .321 | .106 |
| | 3 | 3.129 | .129 | .590 | .364 | 3.107 | .107 | .533 | .296 |
| | 4 | 4.230 | .230 | .910 | .881 | 4.230 | .230 | .853 | .780 |
| | 5 | 5.268 | .268 | 1.240 | 1.610 | 5.286 | .286 | 1.299 | 1.770 |

* Each entry is the result of 1000 simulations.

## Multistage Estimation of k

We suggest that a multistage procedure may be useful in the estimation of k. Suppose five observations are taken at random from the population and the method of moments estimate of k is computed. Then another five observations are added to the sample and the MME of k recomputed. The process of adding five more values to the sample and determining the MME of k continues until the last two estimates differ by less than .05 . Then the last MME of k would be taken as the estimate of k.

Table XV contains the results of the computer simulation for this procedure. Notice that the estimated bias, standard deviation, and MSE of k all tend to be reduced over comparable fixed sample size estimates. The values under the "Stopping Criterion Not Met" column represent the number of times the stopping rule was not satisfied after taking 400 observations. These samples were excluded from the computation of the other quantities in the table.

Since there are times when we might need to stop sampling before meeting the stopping criterion, we considered two truncation rules. We assumed we would take as many as one hundred observations per sample. If we had not stopped, then we would either take the estimate based on the hundred observations or determine the two successive estimates closest together and take the second one of these as our estimate of k. Tables XVI and XVII present the results of 1500 computer simulations for these two procedures. Although inconclusive and contrary to intuition, it appears that the second truncation rule may be the better one.

Further simulations have been conducted on the effects of increasing the number of observations between points where the MME of k are calcu-

TABLE XV

MULTISTAGE ESTIMATION OF k*

| $\mu$ | k | Stopping Criterion Not Met | $\overline{N}$ | $s_{\overline{N}}$ | $\overline{\hat{k}}$ | $\widehat{\text{Bias}}$ | $s_{\hat{k}}$ | $\widehat{\text{MSE}}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 43.28 | 1.08 | 1.12 | .12 | .74 | .56 |
| | 2 | 0 | 77.18 | 2.17 | 2.25 | .25 | 1.40 | 2.04 |
| | 3 | 10 | 110.36 | 3.65 | 3.12 | .12 | 2.02 | 4.09 |
| | 4 | 27 | 125.06 | 4.07 | 3.67 | -.33 | 2.21 | 5.01 |
| | 5 | 65 | 142.63 | 4.82 | 4.33 | -.67 | 2.42 | 6.32 |
| 2 | 1 | 0 | 39.14 | .81 | 1.14 | .14 | .61 | .39 |
| | 2 | 0 | 57.57 | 1.36 | 2.13 | .13 | 1.01 | 1.04 |
| | 3 | 0 | 80.08 | 1.96 | 3.22 | .22 | 1.57 | 2.51 |
| | 4 | 0 | 94.77 | 2.55 | 4.40 | .40 | 3.03 | 9.32 |
| | 5 | 4 | 114.58 | 3.27 | 5.40 | .40 | 3.17 | 10.18 |
| 3 | 1 | 0 | 34.13 | .71 | 1.09 | .09 | .51 | .27 |
| | 2 | 0 | 51.25 | 1.21 | 2.14 | .14 | 1.01 | 1.04 |
| | 3 | 0 | 70.30 | 1.66 | 3.14 | .14 | 1.32 | 1.75 |
| | 4 | 0 | 85.10 | 2.11 | 4.25 | .25 | 2.09 | 4.44 |
| | 5 | 0 | 100.44 | 2.58 | 5.16 | .16 | 2.54 | 6.47 |
| 4 | 1 | 0 | 32.79 | .71 | 1.11 | .11 | .48 | .24 |
| | 2 | 0 | 48.23 | 1.08 | 2.17 | .17 | .93 | .89 |
| | 3 | 0 | 62.71 | 1.46 | 3.22 | .22 | 1.22 | 1.54 |
| | 4 | 0 | 77.83 | 1.92 | 4.23 | .23 | 1.66 | 2.82 |
| | 5 | 0 | 88.85 | 2.20 | 5.05 | .05 | 2.25 | 5.05 |
| 5 | 1 | 0 | 32.48 | .62 | 1.12 | .12 | .51 | .28 |
| | 2 | 0 | 46.29 | 1.02 | 2.15 | .15 | .94 | .90 |
| | 3 | 0 | 60.57 | 1.37 | 3.27 | .27 | 1.26 | 1.67 |
| | 4 | 0 | 69.93 | 1.68 | 4.14 | .14 | 1.80 | 3.26 |
| | 5 | 0 | 86.67 | 2.18 | 5.17 | .17 | 2.65 | 7.07 |

* Each entry is based on 500 simulations.

TABLE XVI

EFFECT OF USING ESTIMATE BASED ON 100 OBSERVATIONS IF STOPPING CRITERION FOR MULTISTAGE ESTIMATION OF k NOT MET*

| $\mu$ | k | Number of Times 100 Observations Taken | $\overline{N}$ | $s_{\overline{N}}$ | $\overline{\hat{k}}$ | $\hat{\text{Bias}}$ | $s_{\hat{k}}$ | $\hat{\text{MSE}}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 58 | 44.960 | .609 | 1.232 | .232 | .811 | .711 |
| | 2 | 368 | 63.663 | .778 | 3.074 | 1.074 | 8.634 | 75.698 |
| | 3 | 624 | 73.703 | .768 | 5.943 | 2.943 | 25.926 | 680.815 |
| | 4 | 765 | 78.987 | .723 | 9.549 | 5.549 | 56.282 | 3198.427 |
| | 5 | 866 | 80.700 | .725 | 9.489 | 4.489 | 46.759 | 2206.542 |
| 2 | 1 | 9 | 37.813 | .488 | 1.166 | .166 | .598 | .385 |
| | 2 | 144 | 55.810 | .689 | 2.280 | .280 | 1.253 | 1.648 |
| | 3 | 369 | 66.607 | .729 | 3.610 | .610 | 4.005 | 16.408 |
| | 4 | 591 | 75.460 | .710 | 5.609 | 1.609 | 10.444 | 111.674 |
| | 5 | 695 | 78.357 | .690 | 7.710 | 2.710 | 17.964 | 330.038 |
| 3 | 1 | 0 | 34.870 | .420 | 1.156 | .156 | .500 | .274 |
| | 2 | 73 | 50.867 | .633 | 2.243 | .243 | .988 | 1.034 |
| | 3 | 249 | 63.010 | .706 | 3.355 | .355 | 1.665 | 2.897 |
| | 4 | 471 | 72.080 | .710 | 4.693 | .693 | 3.664 | 13.903 |
| | 5 | 622 | 76.693 | .704 | 6.392 | 1.392 | 6.911 | 49.696 |
| 4 | 1 | 1 | 33.383 | .405 | 1.139 | .139 | .498 | .267 |
| | 2 | 44 | 48.060 | .606 | 2.229 | .229 | .997 | 1.046 |
| | 3 | 197 | 59.947 | .690 | 3.284 | .284 | 1.517 | 2.381 |
| | 4 | 382 | 69.210 | .711 | 4.447 | .447 | 1.996 | 4.185 |
| | 5 | 540 | 75.220 | .690 | 5.595 | .595 | 2.964 | 9.141 |
| 5 | 1 | 0 | 33.253 | .400 | 1.169 | .169 | .488 | .267 |
| | 2 | 36 | 46.043 | .577 | 2.191 | .191 | .839 | .740 |
| | 3 | 144 | 58.137 | .660 | 3.224 | .224 | 1.248 | 1.608 |
| | 4 | 315 | 65.550 | .716 | 4.364 | .364 | 1.824 | 3.459 |
| | 5 | 455 | 72.023 | .711 | 5.483 | .483 | 2.632 | 7.161 |

* Each entry is based on 1500 simulations.

TABLE XVII

MINIMUM DIFFERENCE TRUNCATION RULE APPLIED IF STOPPING CRITERION OF MULTISTAGE ESTIMATION OF k NOT MET BY 100 OBSERVATIONS*

| $\mu$ | k | Number of Times 100 Observations Taken | $\bar{N}$ | $s_{\bar{N}}$ | $\bar{\hat{k}}$ | $\hat{\text{Bias}}$ | $s_{\hat{k}}$ | $\hat{\text{MSE}}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 47 | 45.393 | .605 | 1.255 | .255 | .857 | .799 |
| | 2 | 362 | 64.593 | .756 | 2.646 | .646 | 2.988 | 9.347 |
| | 3 | 652 | 74.393 | .770 | 5.082 | 2.082 | 27.307 | 749.992 |
| | 4 | 773 | 78.533 | .730 | 5.991 | 1.991 | 23.197 | 542.070 |
| | 5 | 888 | 82.183 | .701 | 6.117 | 1.117 | 13.357 | 179.656 |
| 2 | 1 | 3 | 37.257 | .468 | 1.155 | .155 | .602 | .387 |
| | 2 | 141 | 55.590 | .675 | 2.347 | .347 | 1.494 | 2.352 |
| | 3 | 352 | 66.420 | .728 | 3.642 | .642 | 3.563 | 13.106 |
| | 4 | 621 | 74.870 | .741 | 5.184 | 1.184 | 8.142 | 67.692 |
| | 5 | 741 | 79.860 | .688 | 6.651 | 1.651 | 11.089 | 125.694 |
| 3 | 1 | 0 | 35.183 | .416 | 1.153 | .153 | .494 | .267 |
| | 2 | 80 | 50.783 | .632 | 2.178 | .178 | 1.009 | 1.051 |
| | 3 | 247 | 63.470 | .707 | 3.306 | .306 | 1.696 | 2.969 |
| | 4 | 470 | 71.230 | .723 | 4.673 | .673 | 3.535 | 12.948 |
| | 5 | 586 | 75.610 | .697 | 5.962 | .962 | 6.360 | 41.377 |
| 4 | 1 | 0 | 33.017 | .410 | 1.123 | .123 | .505 | .270 |
| | 2 | 33 | 48.223 | .571 | 2.219 | .219 | .887 | .834 |
| | 3 | 180 | 59.717 | .692 | 3.334 | .334 | 1.611 | 2.706 |
| | 4 | 363 | 67.990 | .718 | 4.385 | .385 | 2.163 | 4.829 |
| | 5 | 524 | 73.653 | .709 | 5.899 | .899 | 4.851 | 24.343 |
| 5 | 1 | 1 | 32.620 | .385 | 1.147 | .147 | .455 | .229 |
| | 2 | 38 | 46.700 | .580 | 2.200 | .200 | .938 | .920 |
| | 3 | 116 | 56.840 | .659 | 3.240 | .240 | 1.355 | 1.893 |
| | 4 | 302 | 66.923 | .691 | 4.404 | .404 | 1.841 | 3.552 |
| | 5 | 488 | 74.250 | .679 | 5.618 | .618 | 2.752 | 7.957 |

* Each entry is based on 1500 simulations.

lated and of requiring a smaller difference in the last two estimates. Both tend to rapidly increase the average sample size while continuing to reduce the mean square error.

# CHAPTER VI

## TESTING AND ESTIMATION OF A COMMON k

In this chapter, we shall determine the maximum likelihood estimator of a k common to several negative binomial populations which may have different means. The likelihood ratio test (LRT) is then developed. Some comparisons in the precision of the estimate of the common k and the power of the tests when using the likelihood procedure and that due to Bliss and Owen (6) are also made. All of the work in this chapter is based on samples of fixed size.

Consider a random sample of size $n_i$ from population i, $i = 1, 2, \ldots, t$, where $\sum_{i=1}^{t} n_i = n$. Let $x_{ib}$ be the b-th observation from the i-th population. Denote the number of observations in population i with the value j by $m_{ij}$. Define $\mu_i$ to be the mean of the i-th population. Assume that k is constant for each population but the means may vary. Then the natural logarithm of the likelihood function is

$$\begin{aligned} \ell n\, L &= \sum_{i=1}^{t} \sum_{j=0}^{\infty} m_{ij}\, \ell n\, P(x_i = j) \\ &= \sum_{i=1}^{t} \sum_{j=0}^{\infty} m_{ij} \Bigg[ \sum_{s=0}^{j-1} \ell n(k_c + s) - \ell n\, j! + k_c\, \ell n\, k_c + j\, \ell n\, \mu_i \\ &\qquad - (k_c + j)\, \ell n(\mu_i + k_c) \Bigg] . \end{aligned}$$

Taking the derivative of L with respect to $\mu_i$, we determine that the MLE of $\mu_i$ is $\overline{X}_{i\cdot}$. Using this fact and differentiating $\ell n\, L$ with respect to $k_c$, we have that the MLE of the common value of k is the root of the following equation in $\hat{k}_c$ :

$$\sum_{i=1}^{t} n_i \,\ell n(\bar{X}_{i\cdot} + \hat{k}_c) - n \,\ell n \,\hat{k}_c = \sum_{i=1}^{t} \sum_{j=0}^{\infty} m_{ij} \sum_{s=0}^{j} \frac{1}{\hat{k}_c + s} .$$

Suppose now we want to test

$$H_0\colon k_1 = k_2 = \ldots = k_t \tag{6.1}$$

versus

$$H_1\colon \text{ not } H_0 .$$

Let $\omega$ be the restricted parameter space under $H_0$ , and $L(\hat{\omega})$ the maximum value of the likelihood function of the sample where the parameters are restricted to $\omega$; hence,

$$L(\hat{\omega}) = \prod_{i=1}^{t} \left\{ \left[ \prod_{b=1}^{n_i} \frac{(\hat{k}_c + x_{ib} - 1)!}{x_{ib}!\hat{k}_c!} \right] \left( \frac{\bar{X}_{i\cdot}}{\bar{X}_{i\cdot} + \hat{k}_c} \right)^{\sum_{b=1}^{n_i} x_{ib}} \left( \frac{\hat{k}_c}{\bar{X}_{i\cdot} + \hat{k}_c} \right)^{n_i k} \right\} .$$

Defining $L(\hat{\Omega})$ to be the maximum value of the likelihood function of the sample where the parameters may take on any value specified in the union of $H_0$ and $H_1$ , we have

$$L(\hat{\Omega}) = \prod_{i=1}^{t} \left\{ \left[ \prod_{b=1}^{n_i} \frac{(\hat{k}_i + x_{ib} - 1)!}{x_{ib}!\hat{k}_i!} \right] \left( \frac{\bar{X}_{i\cdot}}{\bar{X}_{i\cdot} + \hat{k}_i} \right)^{\sum_{b=1}^{n_i} x_{ib}} \left( \frac{\hat{k}_i}{X_{i\cdot} + \hat{k}_i} \right)^{n_i \hat{k}_i} \right\}$$

where $\hat{k}_i$ is the maximum likelihood estimate of k based on the observations from the i-th population.

The likelihood ratio is then denoted by

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} .$$

If $H_0$ is true, -2 times the natural logarithm of $\lambda$ is approximately a $\chi^2$ random variable with t degrees of freedom for large value of n.

We propose to use the maximum likelihood estimate of the common k and to test the hypothesis in (6.1) using the $\chi^2$ approximation of $-2 \ell n \lambda$ under $H_0$ .

Monte Carlo methods were employed to study the properties of our proposed estimation and testing procedure and to compare it with the standard one developed by Bliss and Owen (6). However, we can only present our work in this area as a preliminary step to future endeavors. We believe it gives some insight into the two procedures, but much more must be done before either method is fully understood or a comparison of the two is complete.

In our simulations, we worked with two populations, and we took a sample of fixed size thirty from each one. Both populations had the same mean of one although we did not make this assumption when obtaining our estimates.

We began by running 1,000 simulations based on Bliss and Owen's weighted regression procedure. A sample was drawn from each of the two populations. If either one of them resulted in $s_i^2 < \overline{X}_{i.}$ , it was discarded and a new sample drawn. A weighted regression was performed iteratively until the last two estimates of $\hat{k}_c$ differed by less than .01 . The test for equality of the two k's was based on the two-tailed chi-square test statistic with one degree of freedom. The test was conducted with a stated significance level of .05 .

In Table XVIII, the fraction of the 1,000 simulations in which $H_0$ was rejected is presented for each of the combinations of k's. The diagonal entries represent the cases where the two values of k are equal and $H_0$ is true. In the off-diagonal entries, the k's in the underlying populations are not equal, and $H_0$ is false. In many cases, we reject a true $H_0$ more often than a false one, leading us to believe the test may be biased.

TABLE XVIII

ESTIMATED SIGNIFICANCE LEVEL OF THE WEIGHTED REGRESSION TEST OF A COMMON k

| | | Value of k in Population 1 | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Value of k in Population 2 | 1 | .036 | .038 | .030 | .039 | .028 |
| | 2 | | .043 | .032 | .028 | .035 |
| | 3 | | | .041 | .029 | .044 |
| | 4 | | | | .033 | .049 |
| | 5 | | | | | .048 |

For those cases where we failed to reject $H_0$, we proceeded to obtain an estimate of the common k. As a measure of the precision of the estimate, we computed the estimated bias, the estimated standard deviation, and the estimated MSE. These are presented in Table XIX. Considering the high MSE associated with MME and MLE estimation of k from samples of fixed size, this process seems to give good estimates of the common k.

Next we turned to the proposed estimation and testing procedure based on the likelihood function. Again we ran 1,000 simulations for each parameter combination. The $\chi^2$ approximation is apparently not good for samples of this size since we seldom rejected the null hypothesis. Therefore, we decided to reject $H_0$ if $-2 \ell n \lambda$ was greater than two. The value two was chosen so that the estimated significance levels of the two tests would be about the same.

TABLE XIX

ESTIMATION OF A COMMON k BASED ON THE WEIGHTED REGRESSION PROCEDURE

| $\mu$ | $k_c$ | $\bar{\hat{k}}_c$ | $\widehat{\text{Bias}}$ | $s_{\hat{k}}$ | $\widehat{\text{MSE}}$ |
|---|---|---|---|---|---|
| 1 | 1 | 1.287 | .287 | .981 | 1.045 |
| | 2 | 2.526 | .526 | 2.978 | 9.144 |
| | 3 | 3.528 | .528 | 3.799 | 14.714 |
| | 4 | 4.069 | .069 | 5.308 | 28.182 |
| | 5 | 4.774 | -.226 | 5.483 | 30.117 |

Results of these simulations are presented in Table XX. Although there are some exceptions, we are more likely to reject $H_0$ if it is false than if it is true, indicating this may be an unbiased test. Notice that as we go down the diagonal, the estimated significance level decreases. This is somewhat disturbing since it implies the significance level may depend on the value of the unknown, but equal parameters.

If we failed to reject $H_0$, an estimate of a common k was computed. The estimated bias, estimated standard deviation of the estimate, and the estimated MSE are given in Table XXI. In each case the estimated MSE is less than the corresponding one computed under the Bliss and Owen process.

TABLE XX

ESTIMATED SIGNIFICANCE LEVEL OF THE LIKELIHOOD RATIO TEST OF A COMMON k

| | | Value of k in Population 1 | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Value of k in Population 2 | 1 | .126 | .154 | .194 | .248 | .221 |
| | 2 | | .067 | .086 | .078 | .070 |
| | 3 | | | .059 | .051 | .051 |
| | 4 | | | | .043 | .033 |
| | 5 | | | | | .030 |

From these simulations, there is an indication that the testing procedure based on the likelihood ratio is more powerful than the one based on the weighted regression although neither can be considered "good" for a mean of one and samples of size 30 from each of two populations. Also the maximum likelihood estimates are a little more precise than the regression estimates.

In another attempt to study the comparative power of the two tests, we tried the following approach. A sample of size thirty was drawn from each of two populations having the same values of the parameters $\mu$ and k, and the test statistic for the regression test computed. This was done one thousand times and

$$\alpha = P(T > \chi^2_{calc})$$

estimated for each of the one thousand observed $\chi^2_{calc}$ values, where T represents the possible values of the test statistic. This produced significance levels under the null hypothesis. Then various alternatives were considered. For each alternative, two samples were drawn, one from each population. The test statistic was computed, and the observed significance level, $\hat{F}(\hat{\alpha})$, determined. A plot of $\hat{F}(\hat{\alpha})$ against $\hat{\alpha}$ was made with three alternatives plotted on each graph.

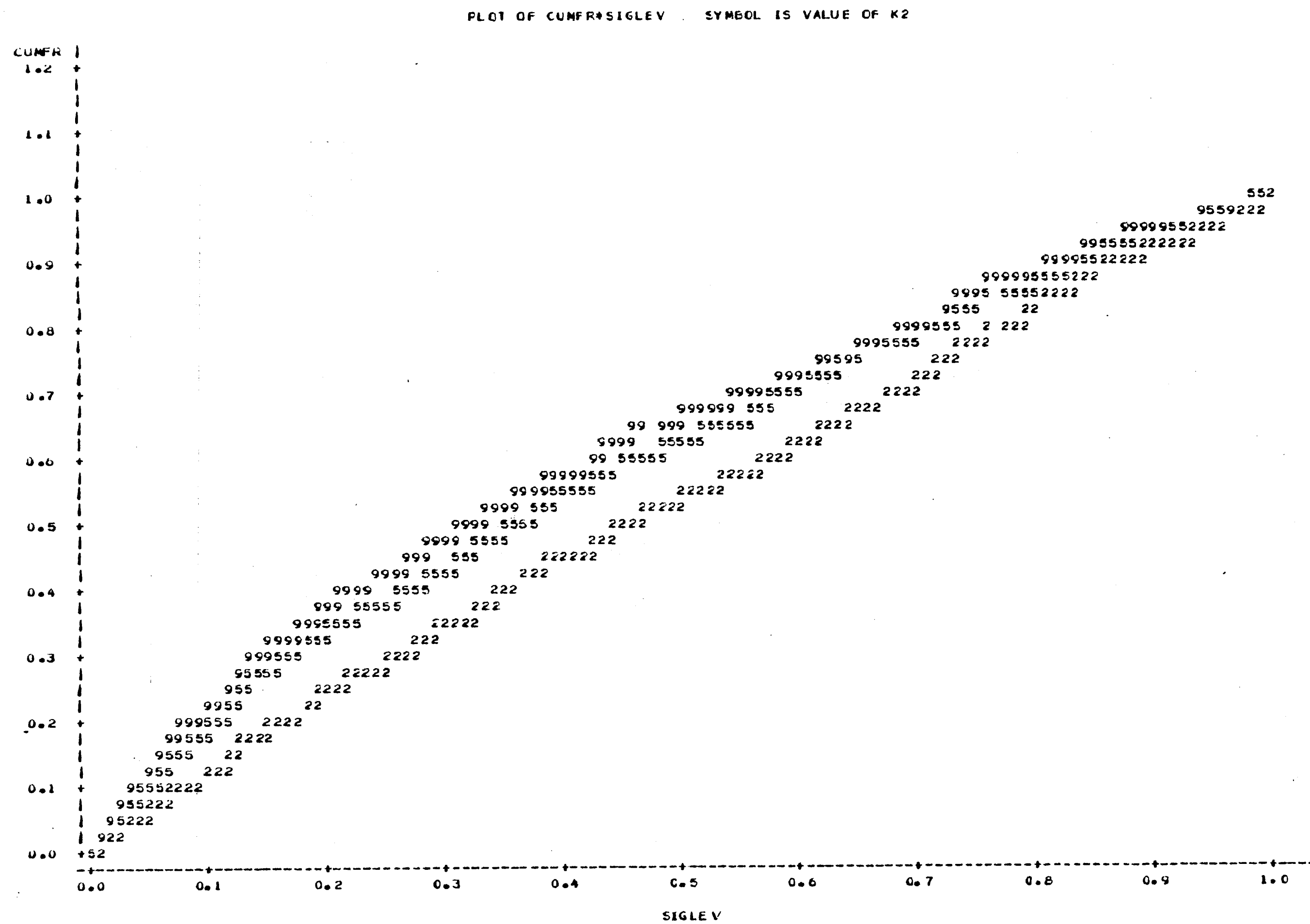
TABLE XXI

MAXIMUM LIKELIHOOD ESTIMATION OF A COMMON k

| $\mu$ | $k_c$ | $\bar{\hat{k}}_c$ | $\hat{\text{Bias}}$ | $s_{\hat{k}}$ | $\hat{\text{MSE}}$ |
|---|---|---|---|---|---|
| 1 | 1 | 1.307 | .307 | .835 | .791 |
| | 2 | 2.544 | .544 | 2.162 | 4.972 |
| | 3 | 3.465 | .465 | 3.565 | 12.927 |
| | 4 | 3.997 | -.003 | 3.609 | 13.025 |
| | 5 | 4.254 | -.746 | 3.645 | 13.845 |

This procedure was repeated for the likelihood ratio test, and the plots based on the two tests compared. Some of these plots are presented in Figures 9 - 16. When the mean is five, both tests are more powerful than when the mean is one. This is reasonable since k affects the shape of the distribution more as $\mu$ increases. The plots of the alternatives under the likelihood ratio are slightly higher than the corresponding ones in the regression procedure, indicating more power.

The plots in Figures 15 and 16 have the significance levels based on $\mu_1 = \mu_2 = 5$ and $k_1 = k_2 = 1$. The alternatives have $\mu_1 = \mu_2 = 5$ and $k_1 = k_2$. We would hope that $\hat{F}(\hat{\alpha}) = \hat{\alpha}$ in these cases. There is a marked tendency for $\hat{F}(\hat{\alpha}) > \hat{\alpha}$ for the regression test statistic, but a rough equality appears to hold for the likelihood ratio test statistic.




NOTE: 2605 OBS HIDDEN

Figure 9. Plot of $\hat{F}(\hat{\alpha})$ of the Weighted Regression Test Statistic Under the Alternatives, $H_1$: $k_1 = 1$, $k_2 = 2$ (2), $H_1$: $k_1 = 1$, $k_2 = 5$ (5), and $H_1$: $k_1 = 1$, $k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0$: $k_1 = 1$, $k_2 = 1$ When $\mu_1 = \mu_2 = 1$

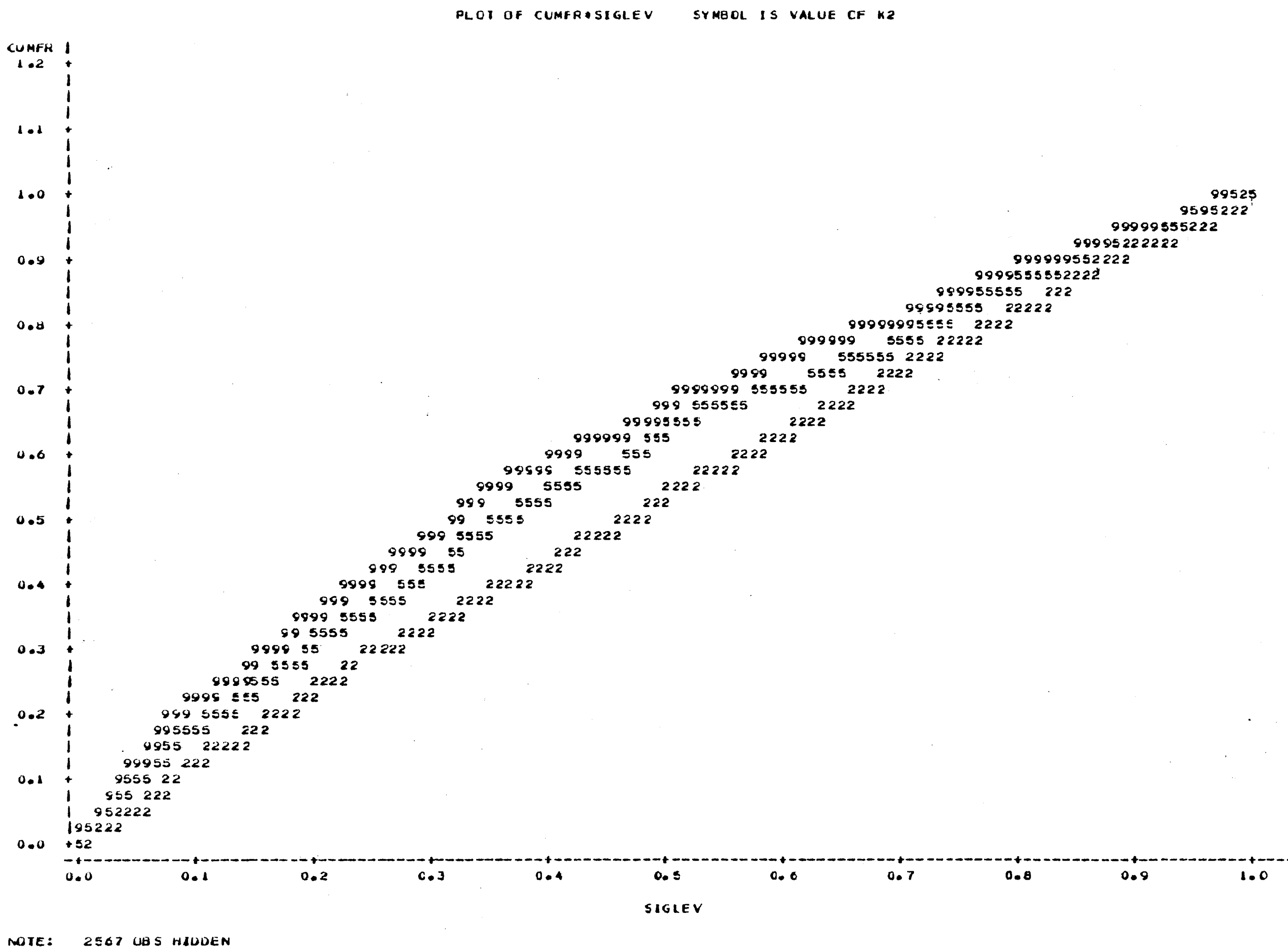



Figure 10. Plot of $\hat{F}(\hat{\alpha})$ of the Likelihood Ratio Test Statistic Under the Alternatives, $H_1$: $k_1 = 1$, $k_2 = 2$ (2), $H_1$: $k_1 = 1$, $k_2 = 5$ (5), and $H_1$: $k_1 = 1$, $k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0$: $k_1 = 1$, $k_2 = 1$ When $\mu_1 = \mu_2 = 1$

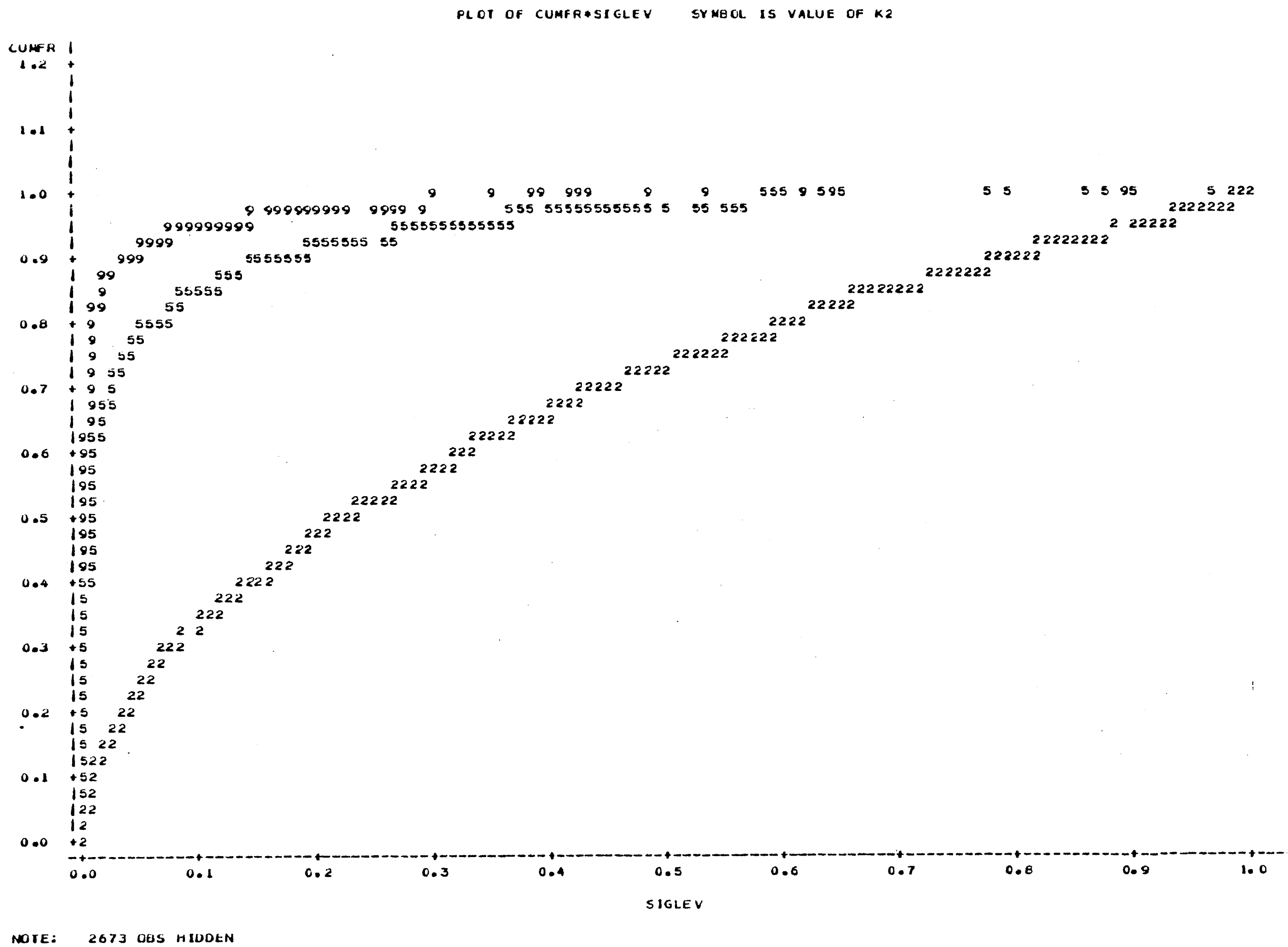



Figure 11. Plot of $\hat{F}(\hat{\alpha})$ of the Weighted Regression Test Statistic Under the Alternatives, $H_1$: $k_1 = 1$, $k_2 = 2$ (2), $H_1$: $k_1 = 1$, $k_2 = 5$ (5), and $H_1$: $k_1 = 1$, $k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0$: $k_1 = 1$, $k_2 = 1$ When $\mu_1 = \mu_2 = 5$

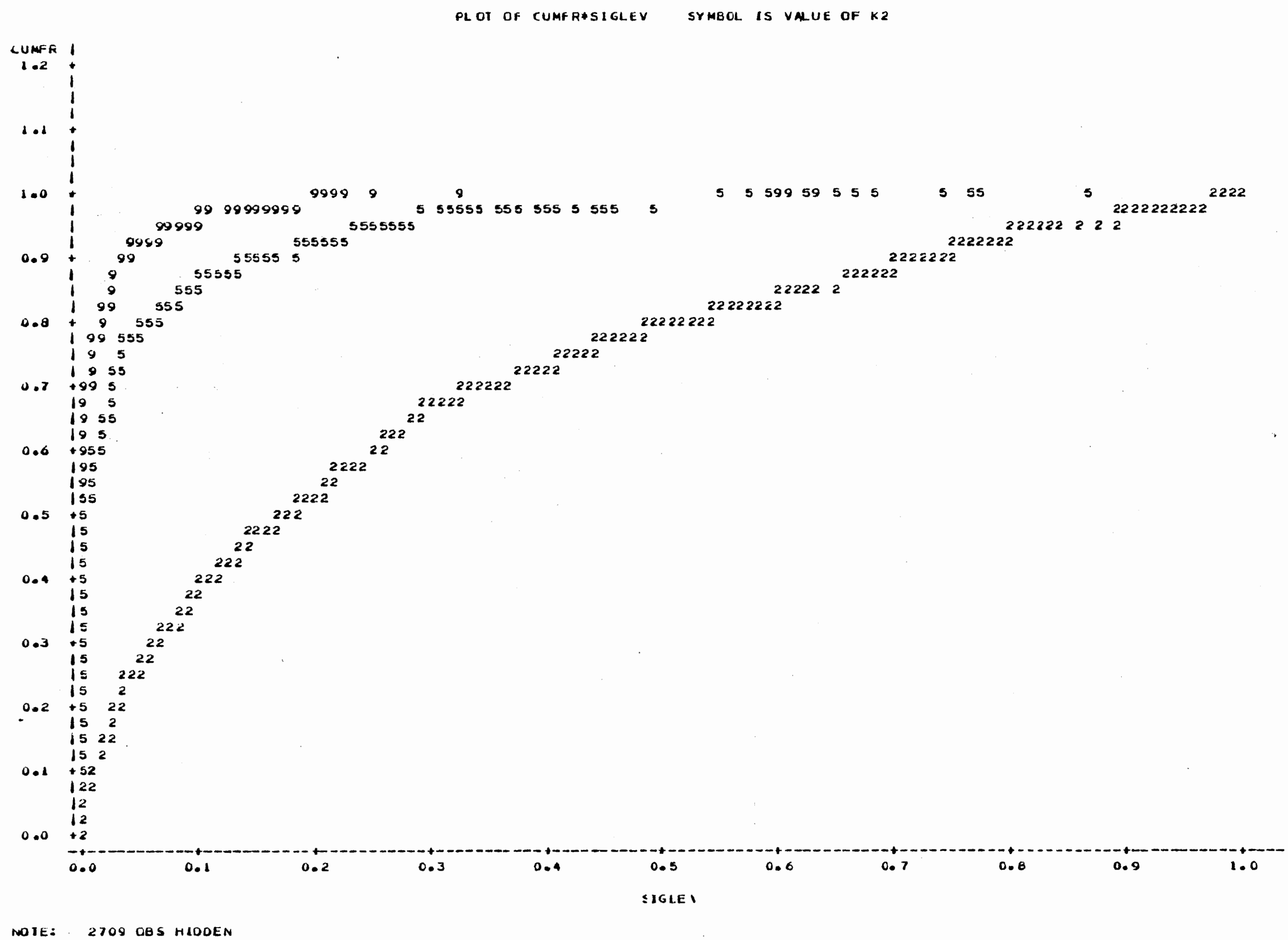
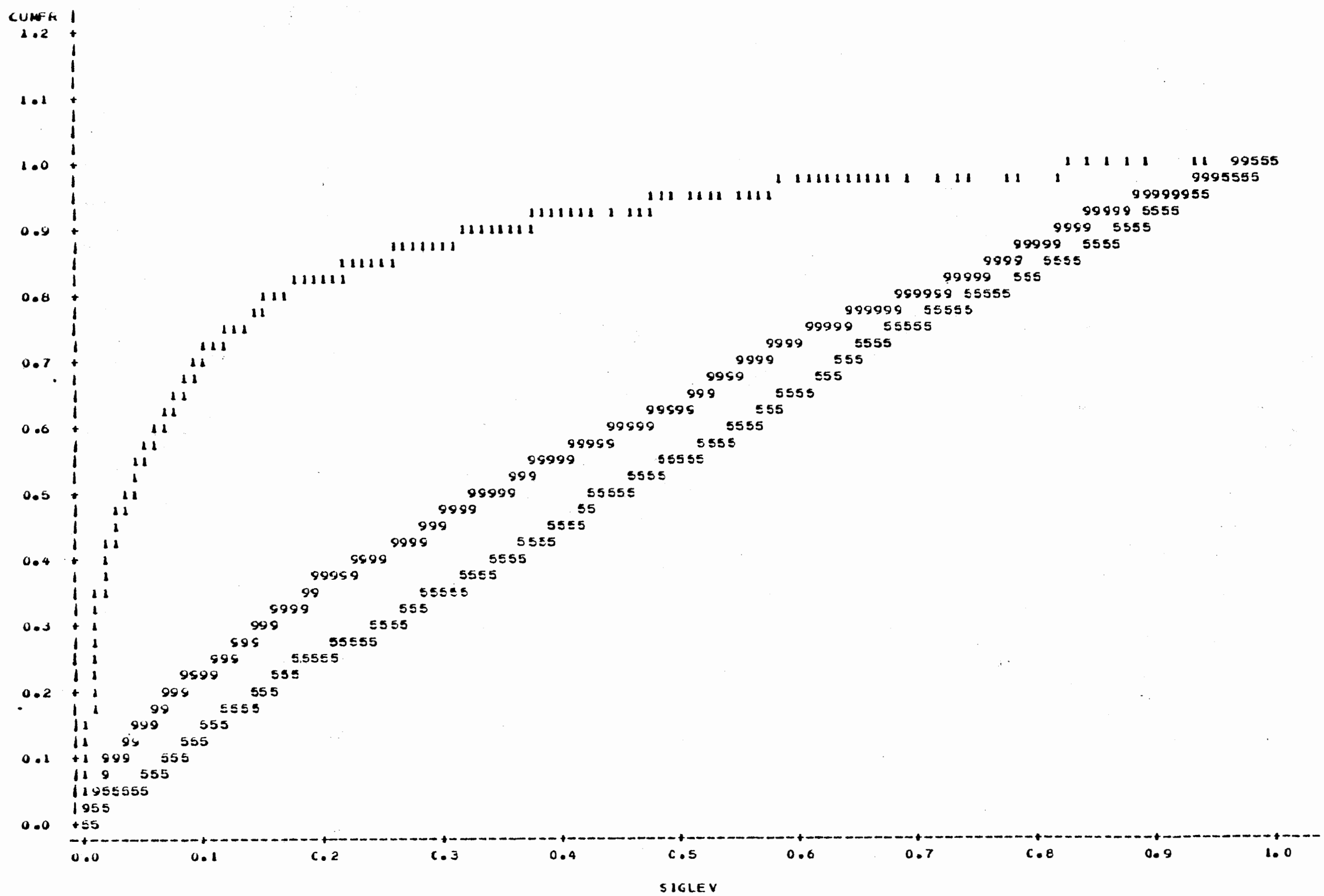



Figure 12. Plot of $\hat{F}(\hat{\alpha})$ of the Likelihood Ratio Test Statistic Under the Alternatives, $H_1$: $k_1 = 1$, $k_2 = 2$ (2), $H_1$: $k_1 = 1$, $k_2 = 5$ (5), and $H_1$: $k_1 = 1$, $k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0$: $k_1 = 1$, $k_2 = 1$ When $\mu_1 = \mu_2 = 5$

PLOT OF CUMFR*SIGLEV    SYMBOL IS VALUE OF K2




NOTE:    2 OBS HAD MISSING VALUES    2575 OBS HIDDEN

Figure 13. Plot of $\hat{F}(\hat{\alpha})$ of the Weighted Regression Test Statistic Under the Alternatives, $H_1$: $k_1 = 4$, $k_2 = 1$ (1), $H_1$: $k_1 = 4$, $k_2 = 5$ (5), and $H_1$: $k_1 = 4$, $k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0$: $k_1 = 4$, $k_2 = 4$ When $\mu_1 = \mu_2 = 5$

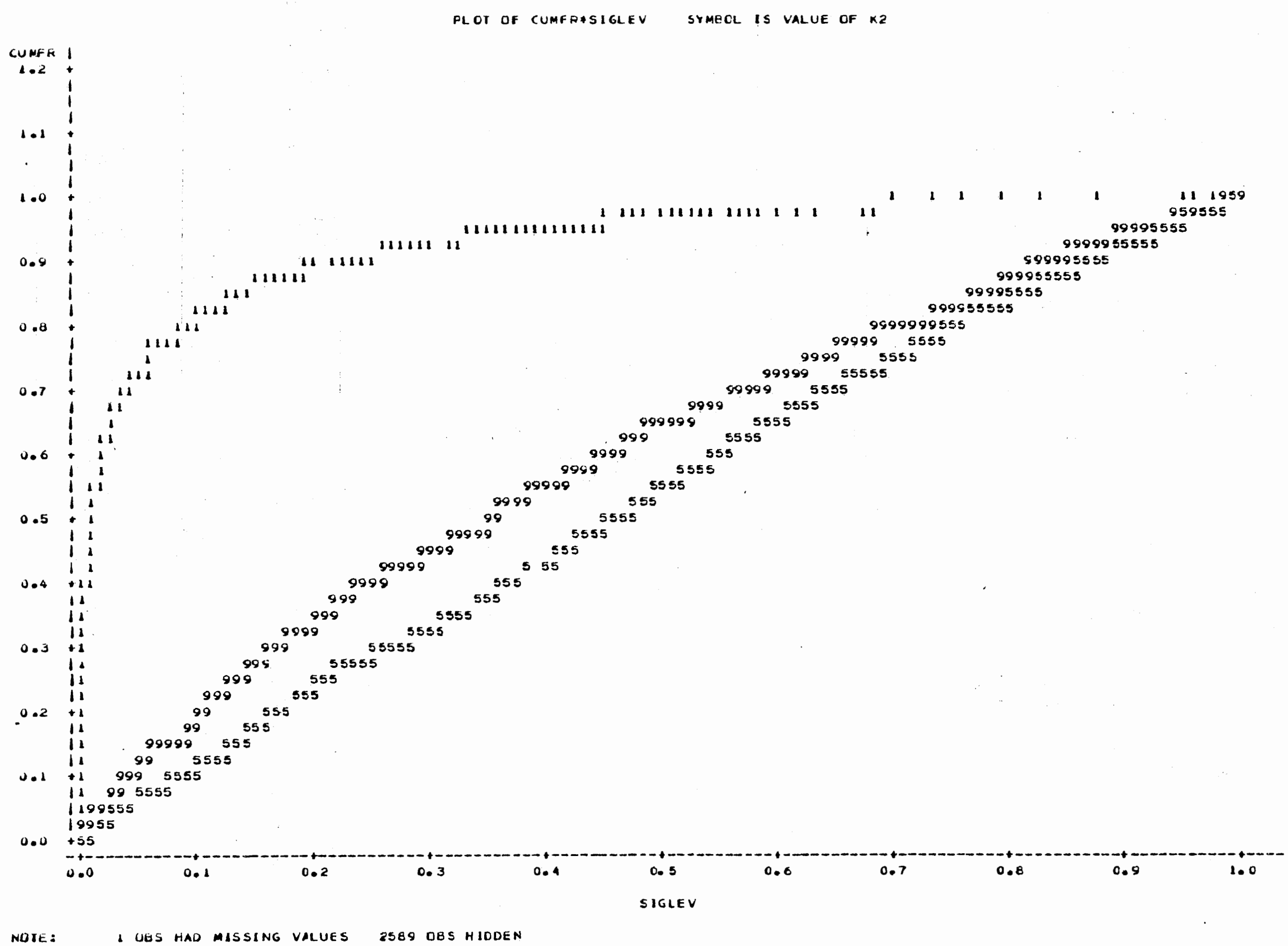



Figure 14. Plot of $\hat{F}(\hat{\alpha})$ of the Likelihood Ratio Test Statistic Under the Alternatives, $H_1: k_1 = 4$, $k_2 = 1$ (1), $H_1: k_1 = 4$, $k_2 = 5$ (5), and $H_1: k_1 = 4$, $k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0$: $k_1 = 4$, $k_2 = 4$ When $\mu_1 = \mu_2 = 5$

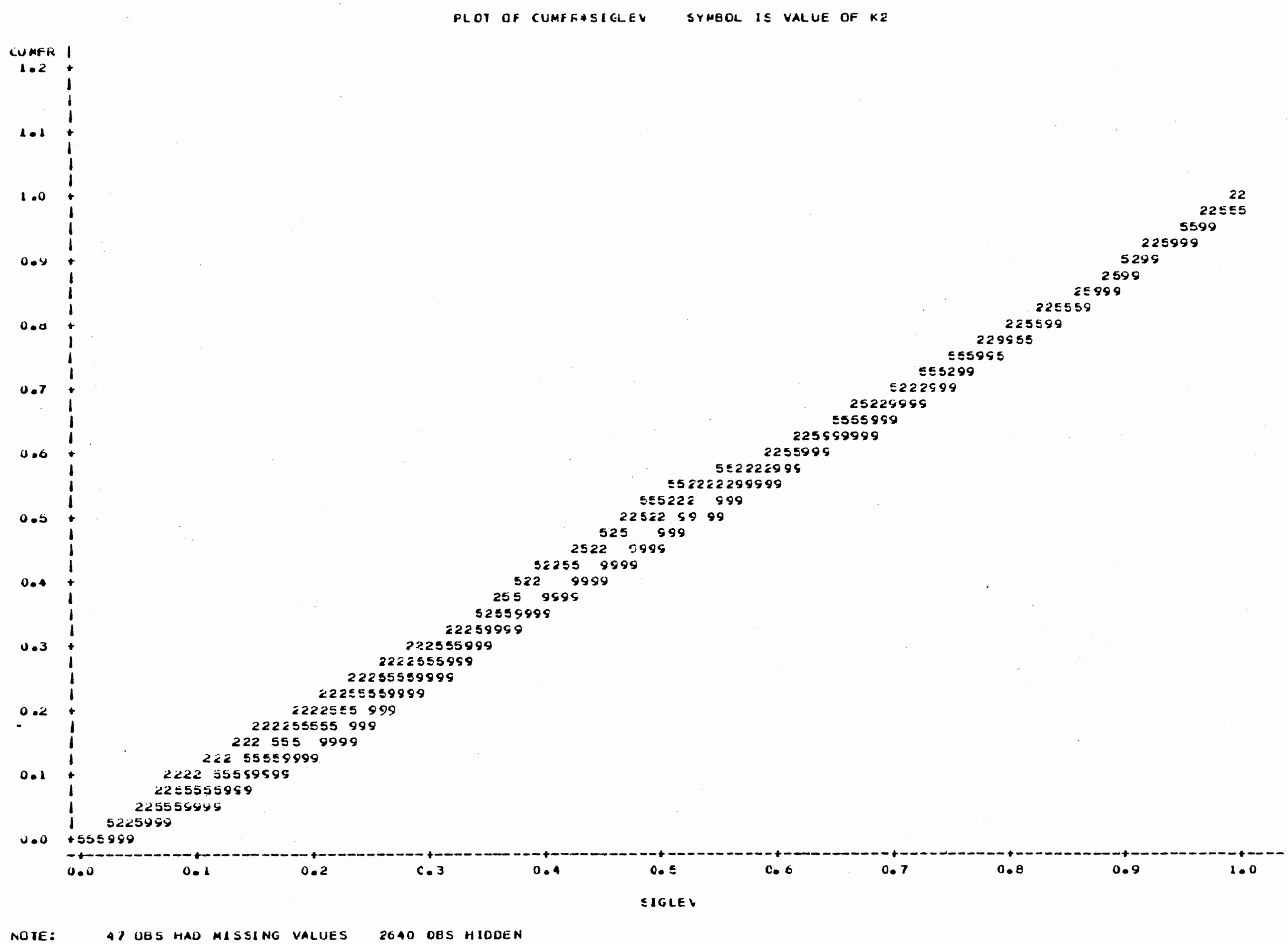



Figure 15. Plot of $\hat{F}(\hat{\alpha})$ of the Weighted Regression Test Statistic Under the Alternatives, $H_1: k_1 = 2$, $k_2 = 2$ (2), $H_1: k_1 = 5$, $k_2 = 5$ (5), and $H_1: k_1 = 9$, $k_2 = 9$ (9), Versus $\alpha$ Under $H_0$: $k_1 = 1$, $k_2 = 1$ When $\mu_1 = \mu_2 = 5$

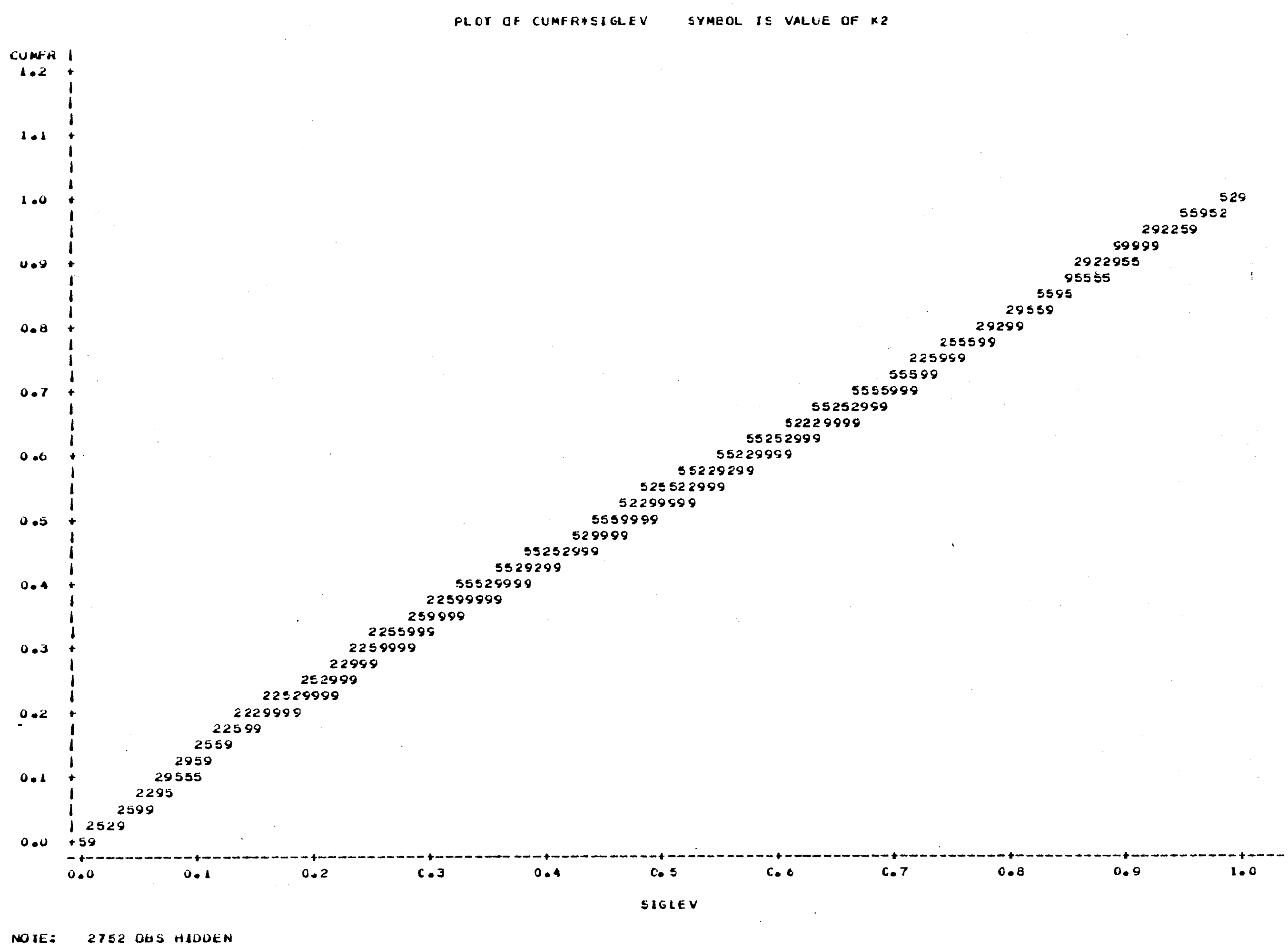



Figure 16. Plot of $\hat{F}(\hat{\alpha})$ of the Likelihood Ratio Test Statistic Under the Alternatives, $H_1: k_1 = 2,\ k_2 = 2$ (2), $H_1: k_1 = 5,\ k_2 = 5$ (5), and $H_1: k_1 = 9,\ k_2 = 9$ (9), Versus $\hat{\alpha}$ Under $H_0: k_1 = 1,\ k_2 = 1$ When $\mu_1 = \mu_2 = 5$

# CHAPTER VII

## SUMMARY

Our study is devoted to the negative binomial distribution. Sequential procedures to estimate the mean with a prescribed degree of precision are developed. A multistage method of estimating the second parameter k is presented and compared to the method of moments and maximum likelihood estimates. Further, a proposed, fixed-sample-size estimation and testing procedure for a k common to several populations with differing means is developed and compared with the standard one.

When presented a sample from a negative binomial distribution with a single unknown parameter $\mu$, we know the minimum variance unbiased estimator of $\mu$ is $\overline{X}$. However, if we want to collect a sample so that the estimate of $\mu$ has a specified degree of precision, then we need to take a sample of optimum fixed size $n^*$. Generally, $n^*$ depends on the unknown parameter and is thus unknown. Three sequential procedures designed to obtain a desired level of precision are developed. One aims at estimating $\mu$ with a specified coefficient of variation of the estimate, C. Another estimates $\mu$ within a proportion p of $\mu$ with confidence $1 - \alpha$, and the last attempts to estimate $\mu$ within d units with confidence $1 - \alpha$. One of the features of each of these methods is that all major computations may be completed before taking the observations, and only the total needs to be considered in deciding whether or not to stop. The limiting behavior of the procedures is investigated, and

Monte Carlo methods are used to study the results for moderate C, p, and d.

A nonparametric, sequential procedure is developed to estimate the mean with a specified coefficient of variation of the estimate, C. In addition to proving some limit results, the behavior for moderate values of C when the negative binomial is the underlying distribution is considered.

For the two-parameter negative binomial distribution, we show that there is no complete sufficient statistic. It is shown by use of simulation that a proposed multistage procedure for estimating k tends to reduce significantly the bias, the standard deviation of the estimate, and consequently the MSE when compared to fixed-sample-size MME and MLE estimates.

Finally, we compared the MLE of a k common to several populations which may have differing means with the estimates obtained by Bliss and Owen's (6) procedure of weighted regression. We also considered the comparative power of the two tests for equality of the k's. Since we only worked with two populations, each having the same mean, and samples of size thirty were drawn from each, we can view this work as only a preliminary step to a more detailed comparison. There are indications, however, that the likelihood-based procedure produces more precise estimates and has greater power.

## A SELECTED BIBLIOGRAPHY


(1) Allen, J., D. Gonzalez, D.V. Gokhale. "Sequential Sampling Plans for the Bollworm, Heliothis zea." Environmental Entomology, 1 (1972), 771-780.

(2) Anscombe, F.J. "The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution." Biometrics, 5 (1949), 165-173.

(3) Anscombe, F.J. "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions." Biometrika, 37 (1950), 358-382.

(4) Anscombe, F.J. "Large Sample Theory of Sequential Estimation." Proceedings of the Cambridge Philosophical Society, 48 (1952), 600-607.

(5) Binns, Michael. "Sequential Estimation of the Mean of a Negative Binomial Distribution." Biometrika, 62 (1975), 433-440.

(6) Bliss, C.I., A.R.C. Owen. "Negative Binomial Distributions with a Common k." Biometrika, 45 (1958), 37-38.

(7) Bortkewitsch. Das Gasetz der kleinen Zahlen. Leipzig: 1898.

(8) Bowman, K.O., L.R. Shenton. Biases of Estimators for the Negative Binomial Distribution. Oak Ridge, Tennessee: Union Carbide Corporation Report No. ORNL-4005, 1966.

(9) Chow, Y.S., Herbert Robbins. "On the Asymptotic Theory of Fixed-Width Confidence Intervals for the Mean." The Annals of Mathematical Statistics, 36 (1965), 457-462.

(10) Chow, Y.S., H. Robbins, H. Teicher. "Moments of Randomly Stopped Sums." The Annals of Mathematical Statistics, 36 (1965), 789-799.

(11) Eggenberger, F., G. Pólya. "Uber die Statistik verketteter Vorgange." Zeitschrift fur angewandte Mathematik und Mechanik, 1 (1923), 279-289.

(12) Feller, W. "On a General Class of 'Contagious' Distributions." The Annals of Mathematical Statistics, 14 (1943), 389-400.

(13) Fisher, R.A. "The Negative Binomial Distribution." Annals of Eugenics, London, 11 (1941), 182-187.

(14) Ghosh, M. , N. Mukhopadhyay. "Sequential Point Estimation of the Mean when the Distribution is Unspecified." Communications in Statistics, A8 (1979), 637-652.

(15) Greenwood, M., G.U. Yule. "An Enquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurence of Multiple Attacks of Disease or Repeated Accidents." Journal of the Royal Statistical Society, Series A, 83 (1920), 255-279.

(16) Gurland, John. "Some Applications of the Negative Binomial and Other Contagious Distributions." American Journal of Public Health, 49 (1957), 1388-1399.

(17) Haldane, J.B.S. "The Fitting of Binomial Distributions." Annals of Eugenics, London, 11 (1941), 179-181.

(18) Johnson, Norman L. and Samuel Kotz. Discrete Distributions. Boston: Houghton Mifflin Company, 1969.

(19) Kuno, Eizi. "A New Method of Sequential Sampling to Obtain Population Estimates with a Fixed Level of Accuracy." Research in Population Ecology, 11 (1969), 127-136.

(20) Kuno, Eizi. "Some Notes on Population Estimation by Sequential Sampling." Research in Population Ecology, 14 (1972), 58-73.

(21) Lehmann, E. L., Henry Scheffé. "Completeness, Similar Regions, and Unbiased Estimation--Part 1." Sankhya, 10 (1950), 304-340.

(22) Lloyd, Monte. "Mean Crowding." Journal of Animal Ecology, 36 (1967), 1-30.

(23) Loeve, M. Probability Theory. Princeton: D. Van Nostrand Company Inc., 1963.

(24) Montmort, P.R. Essai d'analyse sur les jeux de hasards. Paris: 1714.

(25) Mortara. "Sulle variazioni di frequenza di alcuni fenomeni demografici rari." Annali di Statistica, Serie 5, 4, 5-81.

(26) Morris, R.F. "A Sequential Sampling Technique for Spruce Budworm Egg Surveys." Canadian Journal of Zoology, 32 (1954), 302-313.

(27) Mukhopadhyay, N. "Sequential Estimation of Location Parameter in Exponential Distributions." Calcutta Statistical Association Bulletin, 23 (1974), 85-95.

(28) Mukhopadhyay, N. Sequential Point Estimation of the Mean when the Distribution is Unspecified. University of Minnesota Technical Report No. 312, 1978.

(29) Nadás, Arthur. "An Extension of a Theorem of Chow and Robbins on Sequential Confidence Intervals for the Mean." The Annals of Mathematical Statistics, 40 (1969), 667-671.

(30) Oakland, G.B. "An Application of Sequential Analysis to Whitefish Sampling." Biometrics, 6 (1950), 59-67.

(31) Pieters, E.G., C.E. Gates, J.H. Matis, W.L. Sterling. "Small Sample Comparisons of Different Estimators of Negative Binomial Parameters." Biometrics, 33 (1977), 718-723.

(32) Sen, P.K., Malay Ghosh. Sequential Point Estimation of Estimable Parameters Based on U-Statistics. University of North Carolina Institute of Statistics Mimeo Series No. 1236, 1979.

(33) Sproule, R.N. "A Sequential Fixed Width Confidence Interval for the Mean of a U-Statistics." (Unpub. Ph.D. dissertation, University of North Carolina.)

(34) Starr, N. "On the Asymptotic Efficienty of a Sequential Procedure for the Fixed-Width Interval Estimation of the Mean." The Annals of Mathematical Statistics, 37 (1966), 1173-1185.

(35) Starr, Norman, Michael B. Woodroofe. "Remarks on a Stopping Time." Proceedings of the National Academy of Sciences, USA, 61 (1968), 1215-1218.

(36) Student. "On the Error of Counting with a Haemacytometer." Biometrika, 5 (1907), 351-360.

(37) Waters, William E. "A Quantitative Measure of Aggregation in Insects." Journal of Economic Entomology, 52 (1959), 1180-1184.

(38) Whitaker, Lucy. "On the Poisson Law of Large Numbers." Biometrika, 10 (1915), 36-71.

2

# VITA

Linda Jean Willson

Candidate for the Degree of

Doctor of Philosophy

Thesis: ESTIMATION AND TESTING PROCEDURES FOR THE PARAMETERS OF THE NEGATIVE BINOMIAL DISTRIBUTION

Major Field: Statistics

Biographical:

Personal Data: Born in Huntsville, Texas, December 24, 1952, the daughter of Mr. and Mrs. Marvin B. Cornette.

Education: Graduated from White Deer High School, White Deer, Texas, in May, 1971; received Bachelor of Science degree in Mathematics from West Texas State University in 1974; received Master of Science degree in Mathematics from West Texas State University in 1976; completed requirements for the Doctor of Philosophy degree in Statistics at Oklahoma State University in May, 1981.

Professional Experience: Undergraduate assistant, Department of Mathematics, West Texas State University, 1971-1973; graduate teaching assistant, Department of Mathematics, West Texas State University, 1975-1976; graduate teaching assistant, School of Mathematical Sciences, Oklahoma State University, 1977-1980; graduate research assistant, Department of Entomology, Oklahoma State University, summers of 1978, 1979; graduate research assistant, Department of Statistics and Agricultural Experiment Station, Oklahoma State University, 1980-1981.

Professional Organizations: American Statistical Association, The Biometric Society, The Institute of Mathematical Statistics, Sigma Xi.